

**Systematische Verfahrensoptimierung im Bereich
der Mega-Sequenzierung
und ihr exemplarischer Einsatz zur
Analyse des Humangenoms**

Von der Gemeinsamen Naturwissenschaftlichen Fakultät der
Technischen Universität Carolo-Wilhelmina
zu Braunschweig

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigte
D i s s e r t a t i o n

von Dipl. Chemiker Oliver Schön
aus Braunschweig

1. Referent: Prof. Dr. Jürgen Bode

2. Referent: Prof. Dr. Klaus Buchholz

eingereicht am: 29. 08 2002

mündliche Prüfung (Disputation) am: 23.01.2003

2003
(Druckjahr)

Veröffentlichungen und Patente

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Gemeinsamen Naturwissenschaftlichen Fakultät, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Patent:

Picking Tool (Gerät zur Aufnahme und Übertragung biologischer Proben): #198 26 244.2

- Erfinder: O. Schön, B. Neelen, G. Kauer, H. Blöcker
- Anmeldedatum: 15.06.1998
- Internationales Anmeldedatum: 15.06.1999
- Veröffentlichung/Offenlegung: 02.06.1999
- Parallelanmeldung in Europa, Kanada, USA, Japan
- Antrag auf Prüfung wurde gestellt

Publikationen:

The chromosome 21 mapping and sequencing consortium. Hattori *et al.* : The DNA sequence of human chromosome 21. *Nature* 405 (2000), 311-319.

International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 409 (2001), 860-921.

INHALTSVERZEICHNIS.....	I
ABKÜRZUNGSVERZEICHNIS.....	IV
1 EINLEITUNG	1
1.1 Projekte zur Entschlüsselung des Humangenoms	1
1.2 Die Sequenzierungsstrategien zur Bestimmung der Basenabfolge des menschlichen Genoms	5
1.3 Technische Innovationen seit Beginn des Human-Genom-Projekts.....	10
1.4 Zielsetzung der Arbeit.....	12
2 ERGEBNISSE	14
2.1 Optimierungen in der Herstellung von "shotgun"-Banken	14
2.1.1 Untersuchung und Optimierung verschiedener DNA-Präparationsmethoden	14
2.1.1.1 Fällungsverfahren	15
2.1.1.2 Säulenchromatographieverfahren	16
2.1.1.2.1 Qiagen-Maxi-Protokoll (modifiziert)	16
2.1.1.2.2 Macherey & Nagel-Maxi-Kit Variante 1.....	17
2.1.1.2.3 Macherey & Nagel-Maxi-Kit Variante 2.....	18
2.1.1.2.4 Macherey & Nagel-Maxi-Kit Variante 3.....	19
2.1.2 Untersuchung verschiedener Fragmentierungsverfahren	20
2.1.2.1 Ultraschallfragmentierung	21
2.1.2.2 Fragmentieren durch hydrodynamisches Scheren.....	23
2.1.3 Herstellen von DNA-Fragmenten mit glatten Enden und anschließende Fertigstellung von „shotgun“-Banken	25
2.1.4 Zusammenfassung der Optimierungen zur Herstellung von "shotgun"-Banken..	28
2.2 Hybridisierungsexperimente zur Entfernung des Klonierungsvektors aus "shotgun"-Banken	29
2.3 DNA-Matrizengewinnung aus Bakterienzellen mittels PCR zum Einsatz in die Sequenzierreaktion	32
2.4 Sequenzierung und Annotierung von Klonen des Chromosoms 21.....	38
2.5 Entwicklung einer Vorrichtung zur manuellen Übertragung von Zellkolonien in Mikrotiterplatten	47
3 DISKUSSION	52
3.1 Optimierte "shotgun"-Banken-Herstellung.....	52
3.2 Entfernung von Klonierungsvektorsequenz tragenden Bakterienklonen aus "shotgun"-Banken	59
3.3 Sequenzierung von PCR-Produkten	63

3.4	Sequenzierung und Annotation.....	66
3.5	Einsatzmöglichkeiten des Picking-Tools	68
4	MATERIAL UND METHODEN	71
4.1	Geräte	71
4.2	Chemikalien und Enzyme	72
4.3	Materialien	73
4.3.1	Bakterienstämme	73
4.3.2	Plasmide.....	74
4.3.3	Nährmedien und Antibiotika	74
4.3.4	Oligonukleotide	74
4.3.5	Verbrauchsmaterialien.....	76
4.4	Allgemeine Grundtechniken.....	77
4.4.1	Sterilisieren.....	77
4.4.2	Herstellung von Agarplatten.....	77
4.4.3	Herstellung von kompetenten Zellen.....	77
4.4.4	Elektrotransformation von <i>E. coli</i>	78
4.4.5	Lagerung von Bakterienkolonien	78
4.4.6	DNA-Fällung.....	78
4.4.6.1	Ermittlung der Fällungsbedingungen zur PCR-Produktaufreinigung.....	79
4.4.7	Aufreinigung von DNA mit Cäsiumchlorid-Gradienten.....	79
4.4.8	UV-spektrometrische Konzentrationsbestimmung von DNA.....	80
4.5	Isolierung von DNA.....	80
4.5.1	Analytische Isolierung von Plasmiden	80
4.5.2	Isolierung von Plasmiden im 96er-Raster für Sequenzierung	81
4.5.3	Präparative Isolierung von Plasmiden	82
4.5.4	Präparative Isolierung von Cosmid- , BAC- und PAC-DNA.....	83
4.5.4.1	Fällungsverfahren	83
4.5.4.2	Säulenchromatographieverfahren.....	84
4.5.4.2.1	Qiagen Maxi-Kit.....	84
4.5.4.2.2	Macherey & Nagel-Kit Variante 1	85
4.5.4.2.3	Macherey & Nagel-Kit Variante 2	86
4.5.4.2.4	Macherey & Nagel-Kit Variante 3	86
4.5.5	Isolierung von DNA aus Agarosegelen	87
4.6	Modifizierung von DNA.....	88
4.6.1	DNA-Verdau mit Restriktionsenzymen	88
4.6.2	Behandlung von DNA mit ATP-abhängiger DNase	88
4.6.3	Fragmentierung von Cosmid- , BAC- und PAC-DNA.....	88
4.6.3.1	Ultraschallfragmentierung.....	88
4.6.3.2	Fragmentierung durch hydrodynamisches Scheren.....	89
4.6.4	Herstellung glatter DNA-Enden	89
4.6.4.1	Variante mit T4-DNA-Polymerase.....	89
4.6.4.2	Variante mit T4-DNA-Polymerase und Klenow-Fragment	90
4.6.5	Dephosphorylierung	90
4.6.6	Ligation von DNA-Fragmenten.....	91

4.7	Einstellungsparameter für die vollautomatische Übertragung von Zellkolonien mit einem Picking- und Spotting-Roboter.....	91
4.8	Polymerase-Kettenreaktion (PCR)	92
4.8.1	PCR zur DNA-Fragmentgewinnung für die Sequenzierung	92
4.8.2	Herstellung DIG-11-dUTP markierter Sonden für Hybridisierungsexperimente	93
4.8.3	Sequenzierung von DNA.....	94
4.9	Gelelektrophorese.....	97
4.9.1	Agarosegele	97
4.9.2	Polyacrylamidgele für die Sequenzierung	98
4.9.2.1	Sequenzierlauf	99
4.10	Zusammenfügen von „shotgun“-Sequenzdaten mittels eines Assemblierungsprogramms	100
4.11	Hybridisierungsexperimente	101
4.11.1	Herstellung der Filter für Hybridisierungen	101
4.11.2	Hybridisierung	103
4.12	Durchführung der Annotierung.....	104
5	ZUSAMMENFASSUNG	107
6	LITERATURVERZEICHNIS	109
7	ANHANG.....	117
7.1	Anhang A: Elektropherogramme	117
7.2	Anhang B: Verwendete Sequenzierprimer	127
7.2.1	Primer für den Klon 98L15:	127
7.2.2	Primer für den Klon 49E3:	128
7.3	Anhang C: Deletionen und Basenaustausche.....	133
7.3.1	Daten aus dem Vergleich der Klone 98L15 (GBF) und 90B5 (Jena)	133
7.3.1.1	Deletionen.....	133
7.3.2	Daten aus dem Vergleich der Klone 49E3 (GBF) und R73i15 (Jena)	133
7.3.2.1	Deletionen.....	133
7.3.2.2	Einzelbasenaustausche.....	138
7.3.2.3	Mehrfache Basenaustausche.....	143
7.4	Anhang D: Annotierungsdaten.....	144
7.4.1	Graphische Darstellung der Annotierungsergebnisse.....	160

Abkürzungsverzeichnis

Abb.	Abbildung
APS	Ammoniumpersulfat
bp	Basenpaar
ddNTP	β -2',3'-Dideoxynukleosidtriphosphat
DNA	Desoxyribonukleinsäure
dNTP	β -2'-Desoxynukleosidtriphosphat
DTT	Dithiothreitol (threo-1,4-Dimercapto-2,3-butandiol)
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamintetraacetat
EST	Expressed sequence tag
fmol	Femtomol
g	Gramm
h	Stunde
IPTG	Isopropyl- β -thiogalactopyranosid
Kbp	1000 Basenpaare
l	Liter
M	Molar
mg, μ g	Milligramm, Microgramm
min	Minuten
ml, μ l	Milliliter, Microliter
mM	Millimolar
mm, μ m	Millimeter, Micrometer
nM	Nanomolar
nm	Nanometer

Abkürzungsverzeichnis

OD	Optische Dichte
PCR	Polymerase-Ketten-Reaktion
pM	Picomolar
pmol	Picomol
RT	Raumtemperatur
s	Sekunde
SDS	Natriumlaurylsulfat
Taq	<i>Thermus aquaticus</i>
TAE	Tris-Acetate-EDTA-Puffer
TBE	Tris-Borat-EDTA-Puffer
TEMED	N,N,N',N'-Tetramethylethyldiamin
U	Einheit (unit)
UpM	Umdrehungen pro Minute
UV	ultraviolettes Licht
Vol	Volumen
w/v, v/v	Gewicht pro Volumen, Volumen pro Volumen

1 Einleitung

1.1 Projekte zur Entschlüsselung des Humangenoms

Im Zeitraum von 1984 bis 1986 wurde auf mehreren wissenschaftlichen Konferenzen die Idee zur vollständigen Sequenzanalyse des Humangenoms erstmals diskutiert und nahm im Laufe dieser Zeit Formen an (Sinsheimer R.L., 1989; Palca J., 1986).

Als Folge wurde im Jahre 1988 mit dem Human-Genom-Projekt (HGP) das erste internationale wissenschaftliche Bio-Großprojekt ins Leben gerufen, um eine fundierte Grundlage für die Erforschung und das Verständnis des Humangenoms zu schaffen (US National Research Council, 1988). In einem Zeitraum von 15 Jahren und einem Etat von 3 Milliarden US-Dollar sollte das Endziel dieser Initiative, die Entzifferung der gesamten Sequenz des menschlichen Genoms, welches die beachtliche Größe von 3×10^9 Basenpaaren aufweist (Morton, 1991), erreicht sein.

Um die Forschungsschwerpunkte und die Kooperation der beteiligten Arbeitsgruppen zu organisieren, wurde von drei US Behörden (National Institutes of Health NIH, National Center for Human Genome Research NCHGR, Department of Energy Office of Health and Environmental Research DOE) ein 5-Jahresplan aufgestellt (US Department of Health and Human Services, US Department of Energy, 1990). Die anfänglichen Ziele des Projekts waren unter anderem:

- a) die Erstellung genetischer und physikalischer Karten und die Sequenzierung des Humangenoms. Dazu sollten neue genetische Marker (Mikrosatelliten) für eine bessere Auflösung genetischer Karten zur Anwendung kommen (Jeffreys *et al.*, 1985). Ferner war die Definition von sequenzmarkierten Stellen (sequence-tagged sites, STS) als allgemein gültige physikalische Marker ein Kriterium (Olson *et al.*, 1989).
- b) die parallele, in gleicher Weise durchgeführte Untersuchung von Modellorganismen.
- c) die Entwicklung von Software und Datenbanken, die eine Verwaltung und Auswertung solch enormer Datenmengen, wie sie bei der Kartierung und Sequenzierung des Humangenoms anfallen, erst ermöglichen.
- d) der Ausbau von Technologie und Automatisierung der DNA-Sequenzierung.

Bei den Modellorganismen, die im Rahmen des Human-Genom-Projekts untersucht werden sollten, handelt es sich um das Bakterium *Escherischia coli*, die Bäckerhefe *Saccharomyces cerevisiae*, den Fadenwurm *Caenorhabditis elegans*, die Fruchtfliege *Drosophila melanogaster* und die Pflanze *Arabidopsis thaliana*. Die Untersuchungen an Modellorganismen sollten der Etablierung neuer Methoden und Technologien dienen, die sich auf die Sequenzierung des Humangenoms übertragen lassen und so bei der Festlegung der Sequenzierstrategie herangezogen werden konnten. So diente z.B. die Sequenzierung des Genoms des Fadenwurms als experimentelle Ausgangsbasis für die Sequenzierung des menschlichen Genoms.

Eingehende Studien an weniger komplexen und dadurch besser zugänglichen Organismen bilden aber auch die Grundlage für die Erforschung der Genetik und Biologie des Menschen. Die Zusammenhänge von Genotyp und Phänotyp in weniger entwickelten Spezies erlauben wertvolle Rückschlüsse auf die Situation im komplexeren *Homo sapiens*. Die Untersuchungen an Modellorganismen wie z.B. *Drosophila melanogaster* oder *Caenorhabditis elegans* haben zunehmend an Bedeutung bei der Aufklärung der normalen Genregulation, genetisch bedingter Erkrankungen sowie evolutionärer Entwicklungen gewonnen.

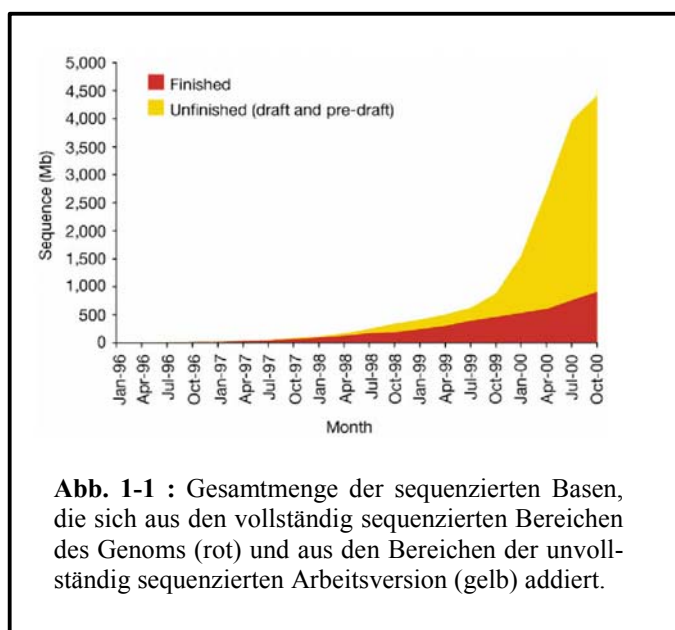
Um ein Projekt solchen Ausmaßes erfolgreich zu gestalten, waren eine detaillierte Organisation der Forschungsvorhaben sowie ein stetiger Austausch über Erfahrungen und Ergebnisse unerlässlich. Kooperation und Kommunikation zwischen den immerhin 20 beteiligten Sequenzierzentren aus den Ländern USA, Japan, England, Deutschland und Frankreich mussten gewährleistet sein. Aus diesem Grund wurde eine internationale Dachorganisation, die *Human Genome Organization* (HUGO) gegründet, deren multidisziplinäre Zentren in den U.S.A hauptsächlich in den DOE Human Genome Centers am Lawrence Berkeley Laboratory (LBL), am Lawrence Livermore National Laboratory (LLNL) und am Los Alamos National Laboratory (LANL) zu finden sind. Entsprechend dem Vorbild in den USA hat man auch in Asien und Europa Programme zur Erforschung des menschlichen Genoms entwickelt. Das weltweite HGP wird mittlerweile von drei HUGO-Zentren aus verwaltet: HUGO America (Bethesda U.S.A.), HUGO Europa (London) und HUGO Pazifik (Tokio).

Da sich die anfänglichen Erfolge schneller als erwartet einstellten, wurde dieser Plan schon nach drei Jahren überarbeitet und die Ziele neu definiert (Collins and Galas, 1993). Hierbei wurden die Schwerpunkte vor allem auf die Erstellung hoch auflösender genetischer Kopplungskarten und physikalischer Karten auf Basis der STS-Marker gelegt. Aber auch die Etablierung neuer Datenverarbeitungstechnologien und Koordination des Informationstransfers zwischen den beteiligten Arbeitsgruppen spielte eine entscheidende Rolle. Darüber hinaus

sollten ethische, juristische und soziale Aspekte des durchaus umstrittenen Projekts berücksichtigt und diskutiert werden.

Der abschließende 5-Jahresplan (Collins F. S. *et al*, 1998) sieht hinsichtlich der Fertigstellung der Sequenz des Humangenoms zwei Meilensteine vor. Natürlich soll die vollständige Sequenz des Humangenoms, mit der Ausnahme von Bereichen des Genoms, die sich mit heutigen Methoden und Technologien nicht klonieren oder sequenzieren lassen, bis zum Ende des Jahres 2003 fertiggestellt werden. Abweichend von der ursprünglichen Strategie, die bearbeiteten Klone gleich vollständig fertig zu stellen, sollte nun bis Ende des Jahres 2001 eine vorläufige Arbeitsversion erstellt werden, die zwar lückenbehaftet ist, aber 90% des Genoms abdeckt. Erst im Anschluss sollte mit dem Schließen vorhandener Sequenzlücken fortgefahren werden. Ein weiterer Punkt dieses Plans sah vor, die erzeugten Daten der unterschiedlichen Forschungseinrichtungen in Datenbanken zu sammeln (z.B. Genome Database, GDB) und diese mit Hilfe des Internets allen Wissenschaftlern vollständig und kostenlos zugänglich zu machen. Ferner galt es auch weiterhin bei der vorhandenen Sequenzieretechnologie den möglichen Durchsatz zu erhöhen und gleichzeitig die Kosten für die Sequenzierung zu verringern.

Für den Strategiewechsel im letzten 5-Jahresplan gab es mehrere Gründe. Einer davon war das stark wachsende Interesse der Wissenschaftsgemeinde an den Sequenzdaten des Humangenoms. Dieses wurde vor allem dadurch hervorgerufen, weil Untersuchungen zeigten, dass eine 90%ige Abdeckung des Genoms mit einer 4-5-fachen Redundanz schon den größten Teil der Informationen in der Sequenz zugänglich machen (Bouck J. *et al*, 1998). Dies ermöglichte Investoren, schon früher als erwartet diese Information zu erhalten und entsprechend für sich nutzbar zu machen. Ein weiterer Grund war das Vorhaben des privaten Biotechnologieunternehmens Celera Genomics mit einer anderen Strategie (auf die unterschiedlichen Strategien wird im Kapitel 1.2 näher eingegangen.) innerhalb von 3 Jahren eine vorläufige Arbeitsversion, die 90% des Humangenoms abdeckt, fertig zu stellen (Venter J. C., 1998). Die Sorge einiger Wissenschaftler, dass die so erhaltene Datenbank vorwiegend zu kommerziellen Zwecken genutzt wird und so nicht allen Firmen und allen Wissenschaftlern frei zur Verfügung steht, hat ebenfalls Einfluss auf den im Rahmen des Human-Genom-Projekts vollzogenen Strategiewechsel gehabt (Marshall E., 1998a; Marshall E., 1998b; Pennisi E., 1999).



Mitte des Jahres 1999, in der Endphase des 5-Jahresplans, konnte mit der Einführung von Sequenzierautomaten mit neuer Technologie (Kapitel 1.3) eine drastische Beschleunigung bei der Erzeugung von Sequenzrohdaten erzielt werden (Abb. 1-1). Waren zum Ende des Pilotprojekts im März 1999 etwa 15% des Genoms sequenziert, so konnte bis Oktober 2000 eine Menge an Rohsequenzen erzeugt werden, die 84% des Genoms 7,5-fach abdeckt.

Ferner konnten in diesem Zeitraum der Euchromatinanteil des Chromosoms 22 annähernd vollständig (Dunham I. *et al.*, 1999) und des Chromosoms 21 vollständig mit einer Genauigkeit von 99,99% sequenziert werden (The chromosome 21 mapping and sequencing consortium, 2000). Im Februar 2001 konnte dann die Arbeitsversion des Humangenoms und die daraus gewonnen Daten veröffentlicht werden (International Human Genome Sequencing Consortium, 2001). Die Daten zum Stand des Sequenzierungsfortschritts sind in der Tabelle 1-1 aufgelistet.

Tab. 1-1: Stand der Sequenzierungsergebnisse der Organismen aus dem Human-Genom-Projekt und der Arbeitsversion des Humangenoms der Fa. Celera Genomics.

Organismus	Jahr	Sequenzierte Basen (Mb)	Vollständigkeit der Abdeckung (%)	Abdeckung des Euchromatins (%)	Vorhergesagte Anzahl von Genen	Anzahl der Gene pro sequenzierter Mb
<i>Saccharomyces cerevisiae</i>	1996	12	93	100	5800	483
<i>Caenorhabditis elegans</i>	1998	97	99	100	19099	197
<i>Drosophila melanogaster</i>	2000	116	64	97	13601	117
<i>Arabidopsis thaliana</i>	2000	115	92	100	25498	221
Human Chromosom 21	2000	34	75	100	225	7
Human Chromosom 22	1999	34	70	97	545	16
Human Genom Arbeitsversion (öffentlich)	2001	2693	84	90	31780	12
Human Genom Arbeitsversion (Celera)	2001	2654	83	88-93	39114	15

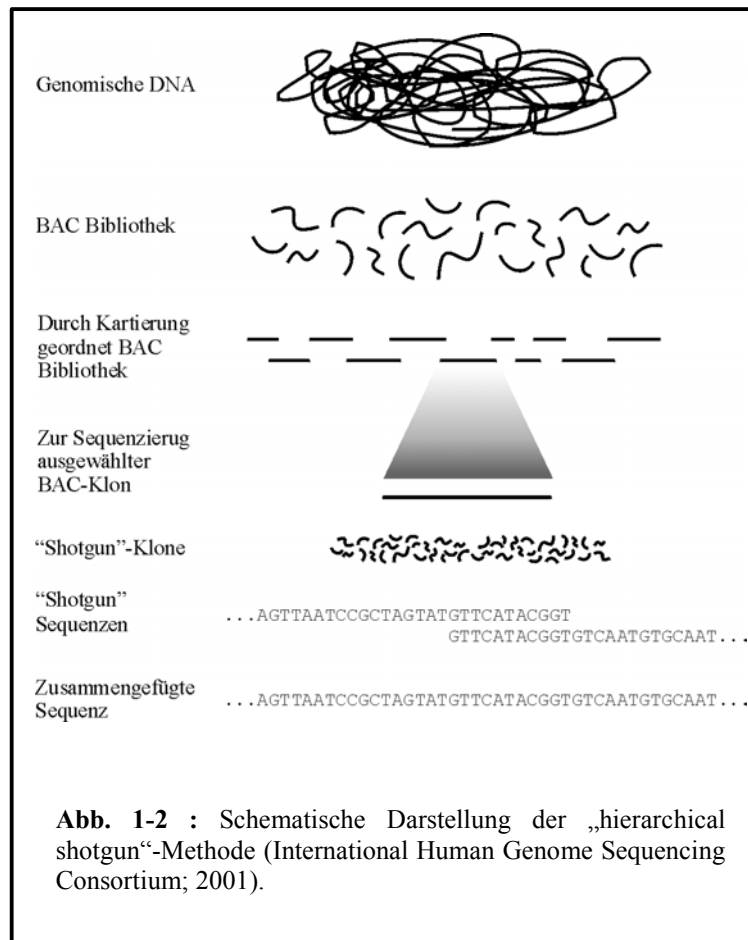
Etwa zum gleichen Zeitpunkt wurden auch die Daten der Arbeitsversion des Genoms veröffentlicht (Tab. 1-1), die von der Fa. Celera Genomics erstellt wurden (Venter J. C. *et al.*, 2001).

1.2 Die Sequenzierungsstrategien zur Bestimmung der Basenabfolge des menschlichen Genoms

Mit der Einführung der sogenannten „shotgun“-Sequenzierung wurde die Grundlage für die Durchführung großer Sequenzierungsprojekte gelegt (Anderson, 1981; Gardner *et al.*, 1981 und Deininger, 1983). Dabei wird die zu sequenzierende DNA mit unterschiedlichen Methoden durch Fragmentierung in 1 bzw. 2-2,5kb große Bruchstücke zerlegt, die sich direkt sequenzieren lassen. Die erhaltenen Sequenzbruchstücke werden anschließend mittels eines Computerprogramms (z. B. GAP4) zusammengefügt, um so die Sequenz der Ausgangs-DNA zu erhalten. Durch optimierte DNA-Fragmentierungsprotokolle konnte eine immer gleichmäßigere Verteilung der DNA-Bruchstücke entlang der Ausgangs-DNA erzielt werden. So konnte die „shotgun“-Methode für die direkte Sequenzierung immer größerer DNA-Moleküle, angefangen bei Plasmiden über BACs und YACs (Wooster R. *et al.*, 1995) bis hin zu Bakteriengenomen (Fleischmann R. D. *et al.*, 1995), erfolgreich eingesetzt werden. Die direkte Sequenzierung ganzer Genome mittels der „shotgun“-Methode wird als „whole-genome shotgun“-Sequenzierung bezeichnet.

Die „whole-genome shotgun“-Methode wurde bisher nur für die Sequenzierung von Genomen eingesetzt, die nur einen geringen Anteil an Sequenzwiederholungen aufweisen, wie z. B. Bakteriengenome (1,5% Sequenzwiederholungen). Im Gegensatz dazu besteht das humane Genom zu mehr als 50% aus Sequenzwiederholungen, die zum Teil sehr lang sind (50-500kb) und hohe Übereinstimmungen (98-99,9%) aufweisen. Die Gefahr, dass die Rohsequenzen durch den verwendeten Assemblierungsalgorithmus fehlerhaft zusammengefügt werden, ist dadurch sehr groß. Ferner kann sich die Behebung solcher Fehler durch die hohe Übereinstimmung der Sequenzwiederholungen als sehr schwierig erweisen. Deshalb stand im Rahmen des Human-Genom-Projekts neben dieser „whole-genome shotgun“-Methode noch eine weitere Methode, die als „hierarchical shotgun“-Sequenzierung bezeichnet wird, für die Anwendung zur Entschlüsselung der Basenabfolge des menschlichen Genoms zur Diskussion. Bei der „hierarchical shotgun“-Methode (Abb. 1-2) wird das Genom, z. B. durch partiellen Verdau mit Restriktionsenzymen, in 100-200kb große Bruchstücke zerlegt, die in ausreichender Anzahl das Genom statistisch abdecken. Diese Bruchstücke werden im Anschluss in geeignete Vektoren ligiert, um so vervielfältigt werden zu können. Die Klone dieser Bibliotheken werden anschließend durch Restriktionsanalysen, sogenannte „finger-prints“, aber auch

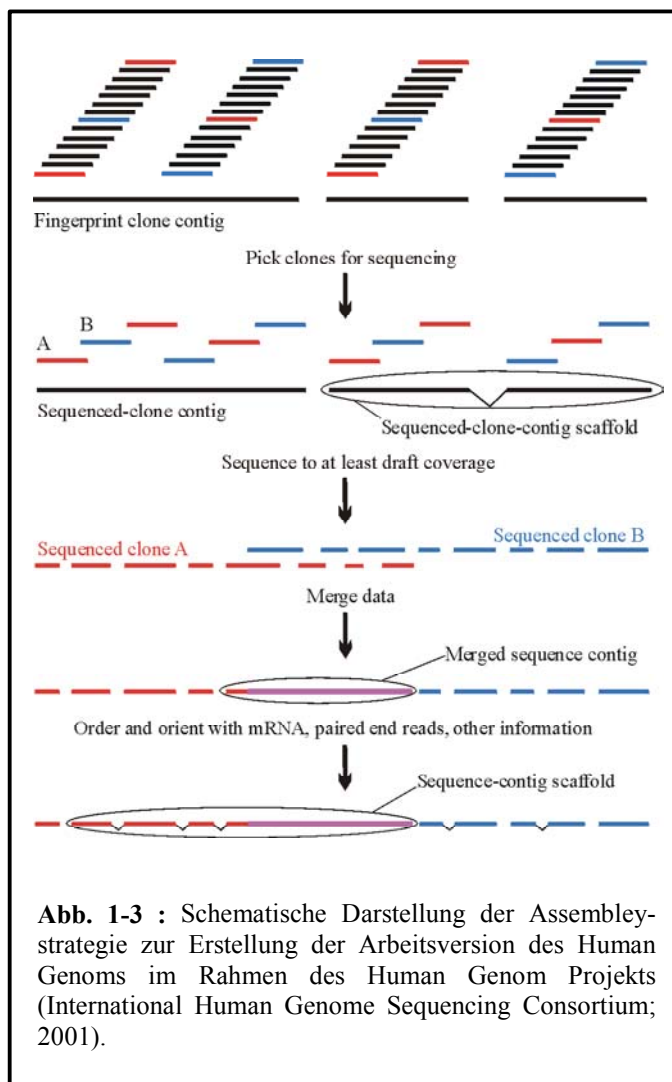
durch Hybridisierungsexperimente entsprechend ihrer Überlappungsbereiche angeordnet. Dabei lässt sich die Lage einzelner Klone, aber auch schon zusammengefügter Contigs, innerhalb des Genoms durch STS-Marker bestimmen. Durch das Anordnen der Klone einer Bibliothek lässt sich das Genom mit einem minimalen Satz entsprechender Klone abdecken, die dann jeweils mittels der „shotgun“-Methode sequenziert werden, um so Klon für Klon die Sequenz des Genoms zu bestimmen. Lücken in der Sequenz dieser Klone, die durch das verwendete DNA-Fragmentierungsverfahren oder durch Klonierungslücken hervorgerufen wurden, können durch direkte



Sequenzierung auf der BAC-DNA mittels entsprechend geplanter Primer („Primer walking“) geschlossen werden.

Wie oben erwähnt war zu Beginn des Human-Genom-Projekts noch kein Genom sequenziert, das einen derart hohen Anteil an Sequenzwiederholungen aufwies. Deshalb wurde letztendlich die „hierarchical shotgun“-Sequenzierung favorisiert. Bei dieser Methode erschien die Wahrscheinlichkeit eines fehlerhaften Zusammenfügens der Rohsequenzen wesentlich geringer und wog somit den erforderlichen Mehraufwand auf, zuvor eine „shotgun“-Bibliothek erstellen und die Klone anordnen zu müssen. Ferner lassen sich bei dieser Methode die weltweit agierenden Sequenzierzentren wesentlich einfacher koordinieren. So lassen sich Klone ausgewählter Bereiche einzelnen Sequenzierzentren zuteilen. Klone können dann dort mit der „shotgun“-Methode sequenziert und eventuelle Lücken in der Sequenz mittels „primer walking“ geschlossen werden können.

Zu Beginn des Human-Genom-Projekts standen deshalb, wie schon erwähnt, die Sequenzierung der Modellorganismen zur Entwicklung und Etablierung neuer Methoden und die Erstellung genetischer und physikalischer Karten des Humangenoms im Mittelpunkt. Zur Erstellung der „shotgun“-Bibliotheken für die Kartierung wurden für die Klonierung großer DNA-Fragmente anfänglich YACs (Burke D. T. *et al.*, 1987) verwendet. Diese wurden in der Folge durch die wesentlich stabileren und dadurch einfacher handhabbaren PACs und dann durch BACs (Shizuya H. *et al.*, 1992) ersetzt. Nach mehrjähriger Vorlaufzeit bei der Kartierung, zu einem Zeitpunkt als die genetische im Gegensatz zur physikalischen Kartierung vollständig abgeschlossen war, wurde dann mit der Sequenzierung des Humangenoms begonnen. Im



weiteren Verlauf des Projekts liefen die Fertigstellung der physikalischen Karte des Genoms und die Sequenzierung parallel. Bei der Sequenzierung der einzelnen Klone wurden diese auch jeweils von den Enden her sequenziert. Diese Endsequenzen konnten wiederum für die Herstellung von Sonden eingesetzt werden, mit denen neben den „fingerprints“ zusätzliche Informationen erhalten werden konnten, um benachbarte Klone zu identifizieren und so die physikalische Karte zu vervollständigen.

Für die Erstellung einer lückenbehafteten Arbeitsversion des Human-genoms musste die Assemblierungs-strategie entsprechend angepasst werden. Die Vorgehensweise ist in der Abb. 1-3 schematisch dargestellt. Be-

reiche der physikalischen Karte, die lückenlos von Klonen abgedeckt sind, werden als „Fingerprint clone contigs“ bezeichnet. Aus diesen Bereichen wurden Klone ausgewählt und sequenziert, die diese bei minimaler Überlappung vollständig abdecken. Bereiche, die durch Zusammenfügen der überlappenden, sequenzierten Klone erhalten werden, bezeichnet man als „Sequenced-clone contigs“. Kann man über Verknüpfungsinformationen die Zusammen-

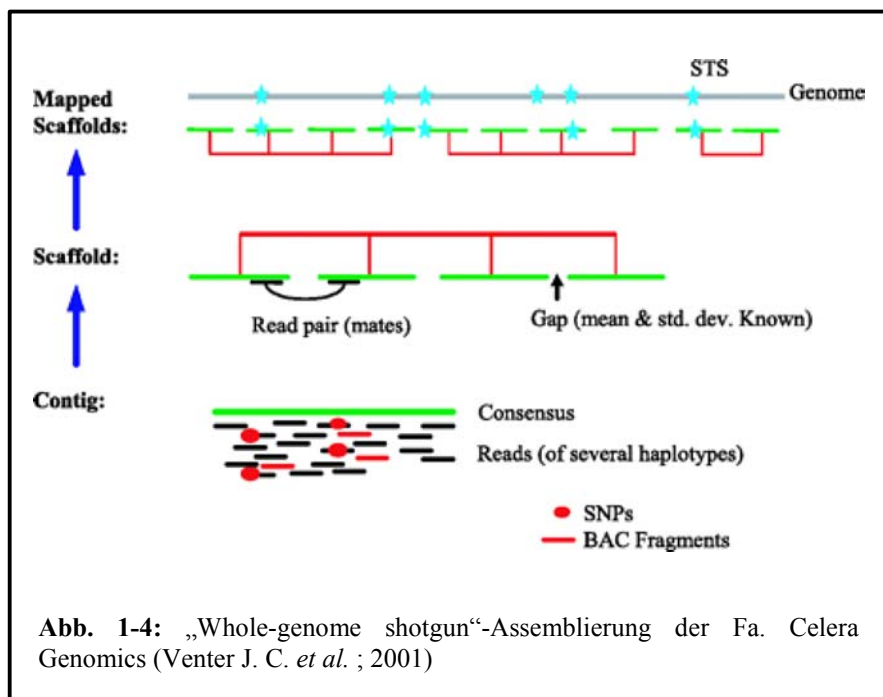
gehörigkeit von „Sequenced-clone contigs“ nachweisen, so wird dieser gesamte Abschnitt als „Sequenced-clone-contig scaffold“ bezeichnet. Gemäß der Strategie des abschließenden 5-Jahresplans wurden die Klone nach der „shotgun“-Sequenzierung assembliert, aber auf ein Schließen der Lücken zwischen den erhalten Contigs wurde vorerst verzichtet. Die weitere Vorgehensweise zur Erstellung einer Arbeitsversion des Humangenoms ist in der Abb.1-3 am Beispiel der Klone A (rot) und B (blau) gezeigt. Aufgrund der Kartierung weiß man, dass es sich um benachbarte Klone handelt, die sich überschneiden. Innerhalb dieses Überschneidungsbereichs können nun, ausgehend von den bekannten, an den Enden der Klone liegenden Contigs überlappende Contigs zusammengefügt werden. Im günstigsten Fall können so alle vorhandenen Lücken im Überschneidungsbereich geschlossen werden. Auf diese Weise lückenfrei erhaltene Sequenzbereiche werden als „Merged sequence contig“ bezeichnet. In der Folge wird versucht die verbliebenen Contigs der Klone in ihrer tatsächlichen Reihenfolge und Orientierung anzuordnen. Dazu werden z. B. Informationen von bekannten mRNA-Sequenzen, ESTs und Sequenzpaaren von „shotgun“-Klonen („paired end reads“), die sich auf einzelnen Contigs wiederfinden lassen, herangezogen. So angeordnete Contigs werden zusammengefasst und als „Sequence-contig scaffold“ bezeichnet.

Die Fa. Celera Genomics hat ihrerseits für die Erstellung einer Arbeitsversion des Humangenoms die „whole-genome shotgun“-Methode favorisiert. Um trotz der oben erwähnten Problematik ein korrektes Zusammenfügen der Rohsequenzen des Humangenoms gewährleisten zu können, sollte ein von „The Institute for Genomic Research“ (TIGR) entwickelter Assemblierungs-Algorithmus verwendet werden. Dieser Algorithmus wurde 1993 zur gleichzeitigen Assemblierung einer sehr großen Anzahl von ESTs entwickelt. Hiermit konnte die Sequenz des 1,8 Mbp großen Genoms von *Haemophilus influenzae* unter Verwendung der „whole-genome shotgun“-Methode entschlüsselt werden (Fleischmann R. D. *et al.*, 1995). Für die Überprüfung der Anwendbarkeit dieser Strategie auf das Human Genom wurde die Sequenzierung des *Drosophila melanogaster* Genoms, einem großen und komplexen eukaryotischen Genom, herangezogen. Der 120 Mbp große Euchromatin-Anteil des Genoms konnte innerhalb eines Jahres sequenziert werden (Adams M. D. *et al.*, 2000; Rubin G. M. *et al.*, 2000; Myers E. W. *et al.*, 2000). Für die Erstellung einer Arbeitsversion dieses Genoms wurde eine Anzahl von Fragmenten sequenziert, die einer 15-fachen Abdeckung des Genoms entspricht. Im Vergleich zur standardmäßigen 7-9-fachen Abdeckung ist dieser Wert annähernd doppelt so hoch. Entsprechend höher waren die Kosten für die Sequenzierung dieses

Genoms. Trotz dieser hohen Abdeckung wurde die Sequenz nicht vollständig erhalten. Die Fertigstellung der Sequenz erfolgte durch „primer walking“ auf einzelnen Klonen.

Für die Bestimmung der Sequenz des Humangenoms sollte innerhalb von 3 Jahren eine Menge von Rohsequenzen erzeugt werden, die einer 10-fachen Abdeckung des Genoms entspricht. Diese Rohsequenzen sollten anschließend zu einer Arbeitsversion des Humangenoms zusammengefügt werden. Eine Schlüsselrolle bei der Assemblierung kam den Verknüpfungsinformationen aus „paired end reads“ zu. Deshalb sollten „shotgun“-Banken mit drei unterschiedlichen Fragmentgrößen (2kbp, 10kbp, 50kbp) hergestellt werden. Für die Herstellung der „shotgun“-Banken wurde von insgesamt fünf Personen DNA aus Blut gewonnen. Von jeder dieser DNA-Proben wurden jeweils, entsprechend den Vorgaben, „shotgun“-Banken hergestellt.

Die Strategieänderung im Rahmen des öffentlichen Human-Genom-Projekts wirkte sich auch auf die Strategie der Fa. „Celera Genomics“ (Venter J. C. *et al.*, 2001) aus. Anstelle der ursprünglich vorgesehen 10-fachen Abdeckung des Genoms wurde im Zeitraum vom 8. September 1999 bis 17. Juni 2000 eine Menge an Rohsequenzen erzeugt, die einer 5,1-fachen Abdeckung entspricht. Wie in der Veröffentlichung (Venter J. C. *et al.*, 2001) beschrieben, wurden zusätzlich Daten entsprechend einer 2,9-fachen Abdeckung von den, in der Datenbank „GenBank“ veröffentlichten Sequenzen des öffentlichen Human-Genom-Projekts hinzugezogen. Die Daten aus der „GenBank“ wurden wieder in Rohsequenzen zerlegt und gingen in dieser Form in die Assemblierung mit ein. Die Assemblierungsstrategie ist in der Abb.



1-4 schematisch dargestellt. Die Rohsequenzen von Celera Genomics (schwarze Linien) als auch die wieder in Rohsequenzen zerlegten Daten (rote Linien) aus dem öffentlichen Human-Genom-Projekt wurden gemeinsam unter Verwendung des TIGR-EST-Algorithmus assembliert. Die

so erhaltenen, zusammenhängenden Sequenzbereiche (Contigs) konnten anschließend durch die Verknüpfungsinformationen der „paired end reads“ zu sogenannten „scaffolds“ angeordnet werden. Durch die Sequenzierung von „shotgun“-Klonen mit den oben erwähnten unterschiedlichen Fragmentgrößen konnten zusätzlich Informationen über die Größen der vorhandenen Lücken zwischen den Contigs erhalten werden. Letztendlich wurden die „scaffolds“ mittels vorhandener Karten des Humangenoms anhand der STS-Markierungen in ihrer Lage auf dem Genom angeordnet, um so die Arbeitsversion des Humangenoms zu erhalten.

Untersuchungen (Waterston R. H. *et al.*, 2001) haben allerdings ergeben, dass in der Veröffentlichung der Fa. „Celera Genomics“ (Venter J. C. *et al.*, 2001) nicht die Analyse des eigenen „whole-genom shotgun“-Datensatzes beschrieben wird. Die Auswahl der Rohsequenzen aus dem HGP-Datensatz erfolgte nicht zufällig. Es wurden gezielt Rohsequenzen ausgewählt, die die gesamte Sequenz des Humangenoms aus dem öffentlichen Human-Genom-Projekt 2,9-fach abdecken. Diese Sequenz wurde aus einer 7,5-fachen Abdeckung erhalten. Computeranalysen haben gezeigt, dass bei dieser Vorgehensweise schon eine 2-fache Abdeckung ausreichend ist, um die vollständige Sequenz durch Assemblierung wiederherstellen zu können. Ferner wurden im Anschluss der Assemblierung dieses gemischten Datensatzes für die Anordnung der erhaltenen Contigs auf dem Humangenom die STS-Markerkarten und die Kartierungsdaten der BAC-Banken aus dem Human-Genom-Projekt verwendet. Deshalb handelt es sich bei den „Celera“-Daten um eine weitere Analyse des HGP-Datensatzes, bei dem zusätzlich die Rohsequenzen aus dem WGS-Datensatz mit in die Assemblierung einbezogen wurden.

1.3 Technische Innovationen seit Beginn des Human-Genom-Projekts

Die Ära der Sequenzierung von DNA begann 1977 mit der von F. Sanger *et al.* entwickelten Methode, die Basenabfolge durch Strangabbruch mittels Didesoxy-Nukleotiden zu ermitteln. Diese enzymatische Reaktion erfolgte in vier getrennten Ansätzen, jeweils für Adenosin (A), Cytosin (C), Guanosin (G) und Thymidin (T) unter zusätzlicher Verwendung von Radioisotopen für die Markierung der entstandenen DNA-Fragmente. Diese Ansätze wurden jeweils in eine Spur auf ein Polyacrylamid-Gel aufgetragen und elektrophoretisch aufgetrennt. Anschließend wurde ein autoradiographisches Bild des Gels erzeugt und die Sequenz anhand dieses Bildes manuell ermittelt. Diese Methode ist sehr sensitiv, aber auch sehr arbeits- und zeitaufwendig.

Die Grundlage für die automatische DNA-Sequenzanalyse wurde mit der Entwicklung fluoreszenzmarkierter Primer gelegt. Auf dieser Basis wurde der erste Sequenzierautomat entwickelt, der eine Echtzeitdetektion der Banden und eine anschließende Sequenzanalyse per

Computer ermöglichte (Smith L. M. *et al.*, 1986). Dabei wurde für jedes Nukleotid ein andersfarbig fluoreszenzmarkierter Primer verwendet. Dadurch musste die Sequenzierreaktion zwar weiterhin in vier getrennten Ansätzen durchgeführt werden, konnte aber anschließend vereinigt und so auf eine Spur des Gels aufgetragen und aufgetrennt werden. Durch die Entwicklung fluoreszenzmarkierter, strangterminierender Didesoxy-Nukleotide (Prober J. M. *et al.*, 1987) konnte auch die Sequenzierreaktion in einem Ansatz durchgeführt werden. In der Folge wurden für diese Vier-Farben-Sequenzierung immer wieder neue Fluoreszenzmarker entwickelt (Ju J. *et al.*, 1995; Metzker M. L. *et al.*, 1996; Rosenblum B. B. *et al.*, 1997; Lee L. G. *et al.*, 1997), mit denen eine ausgeglichene Signalhöhe der einzelnen Nukleotide erzielt werden konnte. Dadurch wurde die automatische Auswertung der erhaltenen Elektropherogramme verbessert.

Aber auch die verwendeten DNA-Polymerasen wurden durch Modifikationen stetig verbessert. Die anfänglich verwendeten modifizierten T7-DNA-Polymerasen (Tabor S. *et al.*, 1988; Tabor S. *et al.*, 1989) wurden später mit der Einführung der „Thermo cycle“-Sequenzierung (Murray V., 1989) durch modifizierte thermostabile DNA-Polymerasen ersetzt, die ebenfalls in der Folge durch Modifikationen weiter verbessert wurden (Reeve M. A. and Fuller C. W., 1995).

Die technologische Entwicklung der Sequenzierautomaten gliedert sich grundsätzlich in zwei unterschiedliche Typen. Einerseits die Sequenzierautomaten, die auf der Ein-Farbstoff-Chemie (ALFexpress, Pharmacia und LONG READIR 4200, LI-COR) beruhen und andererseits die bereits erwähnten Sequenzierautomaten (ABI PRISM™ 377, Applied Biosystems), bei denen die Vier-Farbstoff-Chemie zum Einsatz kommt. Die Entwicklung dieser Sequenzierautomaten verlief unterschiedlich. Geräte wie der ALFexpress oder LICOR waren, durch die Notwendigkeit 4 Taschen des Gels pro Klon belegen zu müssen, von vornherein nicht für die parallele Sequenzierung einer großen Anzahl von Klonen ausgelegt. Deshalb wurde die Entwicklung in Richtung der erzielbaren Leseweiten voran getrieben. Mit Geräten wie dem LONG READIR 4200 können Leseweiten von mehr als 1000 Basen erzielt werden und liegen damit etwa doppelt so hoch wie die eines ABI PRISM™ 377.

Die Entwicklung der Sequenzierautomaten mit Vier-Farbstoff-Chemie ging in Richtung eines größeren Probendurchsatzes. Konnten z. B. mit einem ABI PRISM™ 377 anfänglich 36 Klone gleichzeitig sequenziert werden, so wurde diese Kapazität auf 48, dann 64 und letztendlich 96 Klone gesteigert. Ferner konnte im Bereich der Elektrophorese beobachtet werden, dass sich bei dünneren Gelen die Trennleistung steigert. Es können höhere Spannungen ange-

legt werden und somit die Zeit für die Elektrophorese bei gleichen Leseweiten verkürzt werden.

Diese Miniaturisierung der Elektrophorese wurde mit der Entwicklung der Kapillarelektrophorese-Sequenzierautomaten (Swerdlow H. *et al.*, 1990; Guttman A. *et al.*, 1990; Luckey J. A. *et al.*, 1990)) konsequent fortgesetzt. Aktuelle Geräte dieses Typs sind z. B. der ABI PRISMTM 3700 der Fa. „Applied Biosystems“ oder der MegaBACE 1000 der Fa. „Amersham Biosciences“. Auch hier können 96 Proben parallel aufgetragen werden, aber es können gleiche Leseweiten wie beim ABI PRISMTM 377 in weniger als der Hälfte der Zeit erzielt werden. Ferner verfügen Geräte dieses Typs über die Option der automatischen Probenauftragung und können so ununterbrochen betrieben werden. Dadurch konnte die Sequenzierkapazität auf das drei- bis vierfache gesteigert werden. Bei Geräten dieses Typs verringert sich ebenfalls der Personalaufwand, da die aufwendige Herstellung von Sequenziergelen entfällt.

Die neuesten Geräte dieses Typs erlauben es 384 Proben parallel zu sequenzieren und bieten somit nochmals eine Erhöhung der Sequenzierkapazität um das Vierfache im Vergleich zu den Vorgängermodellen.

1.4 Zielsetzung der Arbeit

Wie einleitend erwähnt, hat sich für die Sequenzierung großer Genome, egal welche der beiden vorgestellten Strategien man betrachtet, die Methode der „shotgun“-Sequenzierung bewährt. Für die erfolgreiche Anwendung dieser Methode ist es notwendig bei der Herstellung von „shotgun“-Banken durch die Art der Fragmentierung eine statistische Verteilung der DNA-Bruchstücke zu gewährleisten. Denn nur so ist es möglich die Ausgangssequenz durch Zusammenfügen der Teilsequenzen annähernd lückenfrei zu erhalten.

Eine besondere Problematik ergibt sich bei der im Rahmen des Human-Genom-Projekts angewendeten, „hierachical shotgun“-Sequenzierstrategie. Die relativ großen Bruchstücke der zu sequenzierenden genomischen DNA befinden sich in geeigneten Vektoren um in *E. coli*-Zellen vervielfältigt werden zu können. Dadurch lässt sich eine ausreichende Menge dieser DNA für die Herstellung von „shotgun“-Banken erhalten. Die Vervielfältigung der DNA in Bakterienzellen hat aber zur Folge, dass die genomische DNA dieser Zellen zu einem mehr oder weniger großen Anteil, ebenso wie der verwendete Vektor in der gesamt präparierten DNA wiederzufinden ist. Die präparierte DNA wird fragmentiert, um daraus „shotgun“-Banken für die Sequenzierung herzustellen. Folglich findet sich auch ein entsprechender Anteil an Klonen in der „shotgun“-Bank, die Bruchstücke der genomischen DNA der *E. coli*-Zellen bzw. des Klonierungsvektors tragen und somit unnützerweise mitsequenziert werden.

Ferner sind mit der fortschreitenden technologischen Entwicklung der Sequenzierautomaten stark ansteigende Sequenzierkapazitäten verbunden (Kapitel 1.3), die aber auch einen steigenden Durchsatz an „shotgun“-Banken erfordern.

Ein Ziel der vorliegenden Arbeit war dementsprechend, die Optimierung der Herstellungsmethode von „shotgun“-Banken. Um eine möglichst vollständige Abtrennung der chromosomalen DNA der *E. coli*-Zellen von der Vektor-DNA zu gewährleisten, sollten Präparationsmethoden untersucht und verbessert werden. Ferner sollten die weiteren, zur Herstellung von „shotgun“-Banken notwendigen biologischen Methoden optimiert und auch neue Gerätschaften etabliert bzw. entwickelt werden, um so dem steigenden Bedarf an „shotgun“-Banken gerecht werden zu können. Da man mit der Wahl der DNA-Präparationsmethode zwar die Abtrennung der chromosomalen DNA der *E. coli*-Zellen beeinflussen kann, diese aber keinen Einfluss auf den Anteil an Bakterienklonen hat, die Bruchstücke des verwendeten Klonierungsvektors tragen, sollten zusätzlich Möglichkeiten aufgezeigt und untersucht werden, die eine nachträgliche, selektive Entfernung dieser Klone aus der „shotgun“-Bank mit vertretbarem Aufwand erlauben.

Zur Zeit erfolgt die DNA-Präparation basierend auf der alkalischen SDS-Lyse (Birnboim, Doly, 1979) mit kommerziellen Kits (z.B. von Qiagen oder Millipore), die es erlaubt DNA von maximal 96 Bakterienklonen parallel zu präparieren. Durch den Zeit- und Arbeitsaufwand dieser Methode und durch die Begrenzung auf ein 96er-Format zeichnet sich auch hier ein Engpass ab, der unter Beibehalt der Methode nur durch die Anschaffung kostspieliger Robotersysteme oder durch einen gesteigerten Personalaufwand zu kompensieren ist. Deshalb stellte die Entwicklung einer alternativen Methode zur DNA-Gewinnung für den Einsatz in die Sequenzierreaktion ein weiteres Ziel der Arbeit dar. Die Methode sollte einen gesteigerten Durchsatz erlauben. Ferner sollte eine parallele Bearbeitung von 384 Bakterienklonen ermöglicht werden, um so eine Anpassung an den Durchsatz aktueller Sequenzierautomaten zu gewährleisten (Kapitel 1.3). Diese Methode sollte so ausgelegt sein, dass eine spätere teilweise oder sogar vollständige automatisierte Durchführung möglich ist.

Ein weiterer Aspekt der vorliegenden Arbeit war die Durchführung vollständiger Sequenzierprojekte. Angefangen bei der Herstellung der „shotgun“-Banken mit anschließender Durchführung der Sequenzierung und Zusammenfügen der Sequenzbruchstücke, bis zur Fertigstellung der lückenfreien Consensussequenz und abschließender Annotierung dieser Sequenz.

2 Ergebnisse

2.1 Optimierungen in der Herstellung von "shotgun"-Banken

Das Prinzip der „hierachical shotgun“-Sequenzierung wurde schon in Kapitel 1.2 ausführlich beschrieben. Um mit diesem Verfahren die stetig steigenden Sequenzierkapazitäten (Kapitel 1.3) nutzen zu können wird eine immer größer werdende Anzahl von "shotgun"-Banken notwendig. Deshalb wurde untersucht, wie die Herstellung von „shotgun“-Banken optimiert werden kann. Es zeigte sich, dass die Effizienz des Verfahrens durch Verbesserung zweier Gesichtspunkte gesteigert werden konnte:

1. Steigerung des Durchsatzes durch Reduktion der Präparationsdauer und Parallelisierung.
2. Steigerung der Zuverlässigkeit des Herstellungsverfahrens.

Die Herstellung von "shotgun"-Banken ist in mehrere Einzelschritte gegliedert. In den nachfolgenden Kapiteln sind die Ergebnisse aus den Optimierungen jedes einzelnen Schritts aufgeführt. Die daraus resultierenden Verbesserungen des Herstellungsverfahrens von „shotgun“-Banken sind in Kapitel 2.1.4 zusammengefasst.

2.1.1 Untersuchung und Optimierung verschiedener DNA-Präparationsmethoden

Alle untersuchten DNA-Präparationsmethoden lysieren zunächst die Zellen durch Einwirken von alkalischer SDS-Lösung (Birnboim, Doly, 1979).

Zur Optimierung wurden verschiedene DNA-Präparationsprotokolle verwendet und hinsichtlich ihrer Effizienz untersucht, mit der die genomische DNA der *E. coli*-Zellen von der Vektor-DNA während der alkalischen SDS-Lyse abgetrennt werden kann. Im Anschluss daran wurde die präparierte DNA weiteren Präparationsschritten unterzogen, um den Anteil genomischer DNA aus den *E. coli*-Zellen an der Gesamtausbeute zu verringern. In einer ersten Versuchsreihe erfolgte die weitere Aufreinigung durch Cäsiumchlorid-Gradienten und einer anschließenden Behandlung mit „Plasmid-Safe“ (Fa. Biozym). Bei „Plasmid-Safe“ handelt es sich um eine ATP-abhängige DNase, die sowohl lineare und zirkuläre einzelsträngige, als auch lineare doppelsträngige DNA abbaut. Zirkuläre, doppelsträngige DNA hingegen bleibt erhalten. Bei einer Präparation im klassischen Stil liegt die genomische DNA aus *E. coli*-Zellen im Gegensatz zur Vektor-DNA fragmentiert vor. Deshalb kann durch die Behandlung

der präparierten DNA mit „Plasmid-Safe“ eine Reduzierung der genomischen DNA erzielt werden. Im Verlauf der experimentellen Arbeiten stellte sich heraus, dass auf die zeitaufwendige Cäsiumchlorid-Gradienten-Zentrifugation gänzlich verzichtet und ausschließlich „Plasmid Safe“ zur Präparation eingesetzt werden konnte.

Die Ergebnisse der einzelnen Präparationsverfahren hinsichtlich der Abtrennung der genomischen DNA aus *E. coli*-Zellen von der Vektor-DNA sind in der Folge jeweils tabellarisch dargestellt. Dort ist die Bezeichnung der Klone angegeben, von denen mit dem jeweiligen Protokoll DNA präpariert und anschließend eine „shotgun“-Bank hergestellt wurde. In einem ersten Schritt wurden von jeder „shotgun“-Bank 96 Klone sequenziert. Trugen mehr als 15% dieser Klone Teile chromosomaler DNA aus *E. coli*-Zellen, so wurde die Bank verworfen und die Herstellung wiederholt. Die Anzahl der Präparationen, die für eine erfolgreiche Herstellung einer „shotgun“-Bank notwendig war, ist ebenfalls angegeben. Bei einer erfolgreich hergestellten Bank ist zusätzlich der Gehalt an chromosomaler DNA aus *E. coli*-Zellen angegeben. Dieser Wert wurde aus den Daten einer vollständig sequenzierten „shotgun“-Bank ermittelt.

2.1.1.1 Fällungsverfahren

Das Fällungsverfahren (Kapitel 4.5.4.1) bietet die Möglichkeit sehr große DNA-Mengen zu präparieren. Bei diesem Verfahren erfolgte die Kultivierung der Bakterienzellen in 3 Liter 2YT-Medium.

Mit diesem Verfahren wurden Ausbeuten von 80-120 µg DNA erzielt. Um die Zellen von dem großen Kulturvolumen in möglichst wenigen Zentrifugationsschritten abtrennen zu können und um eine weitgehende Parallelisierung zu erzielen, wurden jeweils drei 500ml-Zentrifugenbecher pro Kultur eingesetzt. Dadurch konnte die Abtrennung der Zellen in zwei Zentrifugationsschritten erfolgen. Bei sechs Positionen im Rotor bedeutete dies, dass bei einer vorhandenen Zentrifuge allerdings nur zwei Präparationen parallel durchgeführt werden konnten. Der zeitliche Aufwand für eine derartige Präparation mit Cäsiumchlorid-Gradienten (Kapitel 4.4.7) und „Plasmid Safe“-Behandlung (Kapitel 4.6.2) betrug 7 Arbeitstage.

Dieses DNA-Präparationsverfahren wurde bei vier Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Anschließend wurden aus der gewonnenen DNA "shotgun"-Banken hergestellt. Die Ergebnisse hinsichtlich der Eignung des Präparationsverfahrens zur Herstellung von "shotgun"-Banken sind in der Tabelle 2-1 wiedergegeben.

Tab. 2-1: Ergebnisse zum verwendeten Fällungsverfahren. Überschritt der Gehalt an chromosomaler DNA aus *E. coli* bei einer Testsequenzierung von 96 Klonen 15%, so wurde die Herstellung der Bank wiederholt. Bei erfolgreich hergestellten „shotgun“-Banken wurde der angegebene Gehalt an chromosomaler DNA von *E. coli* aus der vollständig sequenzierten Bank ermittelt..

Klonbezeichnung	Insertgröße (kbp)	Anzahl der Präparationen	Erfolgreich hergestellte Bank	Gehalt an chromosomaler DNA aus <i>E. coli</i> (%)
F18108	152,383	1	Ja	15
G1150	178,660	1	Ja	12
111K10	145,656	1	Ja	13,7
73M5	105,999	1	Nein	

Wie aus der Tabelle zu ersehen, wurden aus den Präparationen der 4 Klone von Chromosom 21 drei "shotgun"-Banken hergestellt und in die Sequenzierung eingesetzt. Diese drei Banken wiesen einen durchschnittlichen Gehalt an chromosomaler DNA aus *E. coli* von 13,6% auf. Nur die "shotgun"-Bank des Klons 73M5 wurde verworfen, da der Gehalt an chromosomaler DNA aus *E. coli* den Grenzwert von 15% überschritt. Diese "shotgun"-Bank wurde durch Verwendung eines verbesserten Verfahrens fertiggestellt.

Um einen geringeren Gehalt an chromosomaler DNA aus *E. coli* erzielen zu können und die Präparation zu beschleunigen wurde ein weiteres DNA-Präparationsverfahren entwickelt.

2.1.1.2 Säulenchromatographieverfahren

2.1.1.2.1 Qiagen-Maxi-Protokoll (modifiziert)

In dieser Versuchsreihe wurde der Maxi-DNA-Präparationskit der Fa. Qiagen eingesetzt. Die Durchführung der DNA-Präparation erfolgte wie in Kapitel 4.5.4.2.1 nach einem modifizierten Protokoll. Abweichend vom Herstellerprotokoll wurden 3 Liter 2YT-Medium für die Kultivierung eines Klons eingesetzt. Aufgrund der größeren Zellmenge wurden bei der alkalischen SDS-Lyse 42 ml von jedem Lyse-Puffer eingesetzt. Um die folgende Aufreinigung mittels Anionentauschersäule zu beschleunigen, wurde die so erhaltene DNA-Lösung durch Isopropanol-Fällung um den Faktor 10 aufkonzentriert. Durch die Volumenverringerng konnte die DNA schneller auf die Säule aufgetragen werden.

Zur weiteren Aufreinigung der DNA wurde ein Cäsiumchlorid-Gradient (Kapitel 4.4.7) durchgeführt, auf den eine Behandlung mit „Plasmid-Safe“ (Kapitel 4.6.2) folgte. Wie zuvor beim Fällungsverfahren (Kapitel 2.1.1.1) konnten nur zwei Präparationen parallel durchgeführt werden konnten. Der zeitliche Aufwand für eine derartige Präparation konnte allerdings auf 5 Arbeitstage verkürzt werden.

Dieses DNA-Präparationsverfahren wurde bei fünf Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Anschließend wurden aus der gewonnenen DNA "shotgun"-Banken hergestellt. Die Ergebnisse hinsichtlich der Eignung des Präparationsverfahrens zur Herstellung von "shotgun"-Banken sind in der Tabelle 2-2 wiedergegeben.

Tab. 2-2: Ergebnisse aus der Säulenchromatographie mit dem modifizierten Qiagen-Protokoll. Überschritt der Gehalt an chromosomaler DNA aus *E. coli* bei einer Testsequenzierung von 96 Klonen 15%, so wurde die Herstellung der Bank wiederholt. Bei erfolgreich hergestellten „shotgun“-Banken wurde der angegebene Gehalt an chromosomaler DNA von *E. coli* aus der vollständig sequenzierten Bank ermittelt.

Klonbezeichnung	Insertgröße (kbp)	Anzahl der Präparationen	Erfolgreich hergestellte Bank	Gehalt an chromosomaler DNA aus <i>E. coli</i> (%)
190A9	123,708	1	Ja	1,7
111K10	145,656	1	Nein	
135E14	135,855	1	Nein	
265A22	116,177	1	Nein	
75G13	123,521	1	Nein	

Bei der DNA-Präparation der fünf aufgeführten Klone führte nur die Präparation des Klons 190A9 zu einer verwertbaren "shotgun"-Bank, die allerdings einen sehr geringen Gehalt an chromosomaler DNA aus *E. coli* von 1,7% aufwies. Bei den anderen Klonen wurden die "shotgun"-Banken aufgrund des hohen Gehalts an chromosomaler DNA aus *E. coli* verworfen und die Präparationen mit einem verbesserten Verfahren wiederholt.

2.1.1.2.2 Macherey & Nagel-Maxi-Kit Variante 1

Die Durchführung dieser Präparationsmethode erfolgte wie zuvor mit dem Qiagen-Maxi-Kit. Das Verfahren unterschied sich nur durch die Verwendung der Puffer und der Anionentauschersäule des Macherey & Nagel-Kits (Kapitel 4.5.4.2.2). Ferner wurde für die weitere Aufreinigung der DNA auf den Cäsiumchlorid-Gradienten verzichtet und nur die „Plasmid-Safe“-Behandlung (Kapitel 4.6.2) durchgeführt. Das große Kulturvolumen von 3 Litern bedingte nach wie vor, dass nur von zwei Klonen parallel DNA präpariert werden konnte. Der zeitliche Aufwand für eine derartige Präparation konnte allerdings auf 3 Arbeitstage verkürzt werden.

Dieses DNA-Präparationsverfahren wurde bei sechs Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Anschließend wurden aus der gewonnenen DNA "shot-

gun"-Banken hergestellt. Die Ergebnisse hinsichtlich der Eignung des Präparationsverfahrens zur Herstellung von "shotgun"-Banken sind in der Tabelle 2-3 wiedergegeben.

Tab. 2-3: Ergebnisse aus der Säulenchromatographie mit dem Macherey & Nagel-Kit, Variante 1. Überschritt der Gehalt an chromosomaler DNA aus *E. coli* bei einer Testsequenzierung von 96 Klonen 15%, so wurde die Herstellung der Bank wiederholt. Bei erfolgreich hergestellten „shotgun“-Banken wurde der angegebene Gehalt an chromosomaler DNA von *E. coli* aus der vollständig sequenzierten Bank ermittelt.

Klonbezeichnung	Insertgröße (kbp)	Anzahl der Präparationen	Erfolgreich hergestellte Bank	Gehalt an chromosomaler DNA aus <i>E. coli</i> (%)
73M5	105,999	1	Ja	0,9
135E14	135,855	2	Ja	5,9
265A22	116,177	1	Ja	4,7
75G13	123,521	1	Ja	8,0
M17648	154,894	2	Ja	13,2
E24775	145,250	1	Nein	

Wie aus der Tabelle ersichtlich konnten von fünf der sechs Klone erfolgreich "shotgun"-Banken hergestellt werden. Bei zwei Klonen wurde die Präparation wiederholt, da beim ersten Versuch der Gehalt an chromosomaler DNA aus *E. coli* den Grenzwert von 15% überschritt. Der durchschnittliche Gehalt an chromosomaler DNA aus *E. coli* der vollständig sequenzierten Banken betrug 6,5%. Die Präparation des Klons E24775 wurde mit einem verbesserten Verfahren wiederholt.

2.1.1.2.3 Macherey & Nagel-Maxi-Kit Variante 2

Im Vergleich zur Variante 1 (Kapitel 2.1.1.2.2) wurde das Kulturvolumen auf 500ml 2YT-Medium reduziert und für die alkalische SDS-Lyse Puffervolumen laut Herstellerangaben verwendet (Kapitel 4.5.4.2.3). Im Anschluss an die Präparation wurde die DNA mit „Plasmid-Safe“ (Kapitel 4.6.2) behandelt. Unter diesen Versuchsbedingungen betrug der zeitliche Aufwand für eine derartige Präparation nach wie vor 3 Arbeitstage. Allerdings konnten durch die Verringerung des Kulturvolumens bis zu sechs Klone parallel bearbeitet werden.

Dieses DNA-Präparationsverfahren wurde bei zwei Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Anschließend wurden aus der gewonnenen DNA "shotgun"-Banken hergestellt. Die Ergebnisse hinsichtlich der Eignung des Präparationsverfahrens zur Herstellung von "shotgun"-Banken sind in der Tabelle 2-4 wiedergegeben.

Tab. 2-4: Ergebnisse aus der Säulenchromatographie mit dem Macherey & Nagel-Kit, Variante 2. Überschrift der Gehalt an chromosomaler DNA aus *E. coli* bei einer Testsequenzierung von 96 Klonen 15%, so wurde die Herstellung der Bank wiederholt. Bei erfolgreich hergestellten „shotgun“-Banken wurde der angegebene Gehalt an chromosomaler DNA von *E. coli* aus der vollständig sequenzierten Bank ermittelt.

Klonbezeichnung	Insertgröße (kbp)	Anzahl der Präparationen	Erfolgreich hergestellte Bank	Gehalt an chromosomaler DNA aus <i>E. coli</i> (%)
E24775	145,250	3	Ja	3
11L15	38,896	1	Nein	>15

Vom Klon E24775 wurde nach dreimaliger Wiederholung der DNA-Präparation erfolgreich eine "shotgun"-Bank hergestellt. Diese wies einen Gehalt an chromosomaler DNA aus *E. coli* von 3% auf. Die DNA-Präparation vom Klon 11L15 wurde nicht mit diesem Verfahren wiederholt, da zu diesem Zeitpunkt ein Wechsel des DNA-Fragmentierungsverfahrens erfolgte. Anstelle der bisher durchgeführten Ultraschallfragmentierung (Kapitel 2.1.2.1) erfolgte die Fragmentierung der DNA nun durch hydrodynamisches Scheren (Kapitel 2.1.2.2). Bei diesem Verfahren wird eine geringere DNA-Menge benötigt, um ausreichend Fragmente des gewünschten Längenbereichs zu erhalten. Deshalb konnte das DNA-Präparationsverfahren weiter verbessert werden. Die Präparation des Klons 11L15 wurde mit diesem Verfahren (Kapitel 2.1.1.2.4) wiederholt.

2.1.1.2.4 Macherey & Nagel-Maxi-Kit Variante 3

Da mit dem Wechsel des Fragmentierungsverfahrens (Kapitel 2.1.2.2) eine geringere DNA-Menge zur Herstellung von „shotgun“-Banken benötigt wurde konnte das Kulturvolumen auf 150 ml 2YT-Medium reduziert werden. Ferner wurde das Volumen der Puffer, die bei der alkalischen SDS-Lyse eingesetzt wurden (Kapitel 4.5.4.2.4), von 24 ml auf 30 ml erhöht. Dadurch wurde eine verbesserte Präzipitation der genomischen DNA der *E. coli*-Zellen erzielt. Zusätzlich wies der klare Überstand aus der alkalischen SDS-Lyse im Vergleich zu den vorherigen Präparationsverfahren eine deutlich geringere Viskosität auf. Deshalb konnte die säulenchromatographische Aufreinigung der DNA deutlich schneller erfolgen. Auch bei dieser Variante wurde die DNA mit „Plasmid-Safe“ (Kapitel 4.6.2) nachbehandelt.

Wie zuvor (Kapitel 4.1.1.2.3) konnten maximal sechs Klone parallel bearbeitet werden. Der zeitliche Aufwand für eine derartige Präparation konnte allerdings auf 2 Arbeitstage verkürzt werden.

Dieses DNA-Präparationsverfahren wurde bei 18 Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Anschließend wurden aus der gewonnenen DNA

"shotgun"-Banken hergestellt. Die Ergebnisse hinsichtlich der Eignung des Präparationsverfahrens zur Herstellung von "shotgun"-Banken sind in der Tabelle 2-5 wiedergegeben.

Tab. 2-5: Ergebnisse aus der Säulenchromatographie mit dem Macherey & Nagel-Kit, Variante 3. Überschrift der Gehalt an chromosomaler DNA aus *E. coli* bei einer Testsequenzierung von 96 Klonen 15%, so wurde die Herstellung der Bank wiederholt. Bei erfolgreich hergestellten „shotgun“-Banken wurde der angegebene Gehalt an chromosomaler DNA von *E. coli* aus der vollständig sequenzierten Bank ermittelt.

Klonbezeichnung	Insertgröße (kbp)	Anzahl der Präparationen	Erfolgreich hergestellte Bank	Gehalt an chromosomaler DNA aus <i>E. coli</i> (%)
2A11	46,007	1	Ja	6,5
11L15	38,896	1	Ja	0,2
3D19	43,000	2	Ja	1,1
9B17	35,213	1	Ja	0,4
16M16	41,598	1	Ja	0,7
B7L1C4	106,710	1	Ja	3,4
B15L0C0	80,692	1	Ja	3,3
B1L1C6	41,185	1	Ja	0,3
B11L7C8	100,133	1	Ja	1,8
B6L1C6	30,409	1	Ja	0,7
16C2	111,648	1	Ja	3,5
29H4	107,469	1	Ja	3,6
214G14	168,083	1	Ja	4,3
256M13	73,465	1	Ja	0,3
82L10	112,597	1	Ja	9,3
98L15	116,189	1	Ja	6,0
B62L20	119,182	1	Ja	4,2
49E3	105,570	8	Ja	6,3

Bei allen bearbeiteten Klonen führte diese Präparationsmethode zur erfolgreichen Herstellung einer "shotgun"-Bank. Nur bei zwei Klonen waren dazu Präparationswiederholungen notwendig. Der durchschnittliche Gehalt an chromosomaler DNA aus *E. coli* in den „shotgun“-Banken betrug 3,1%.

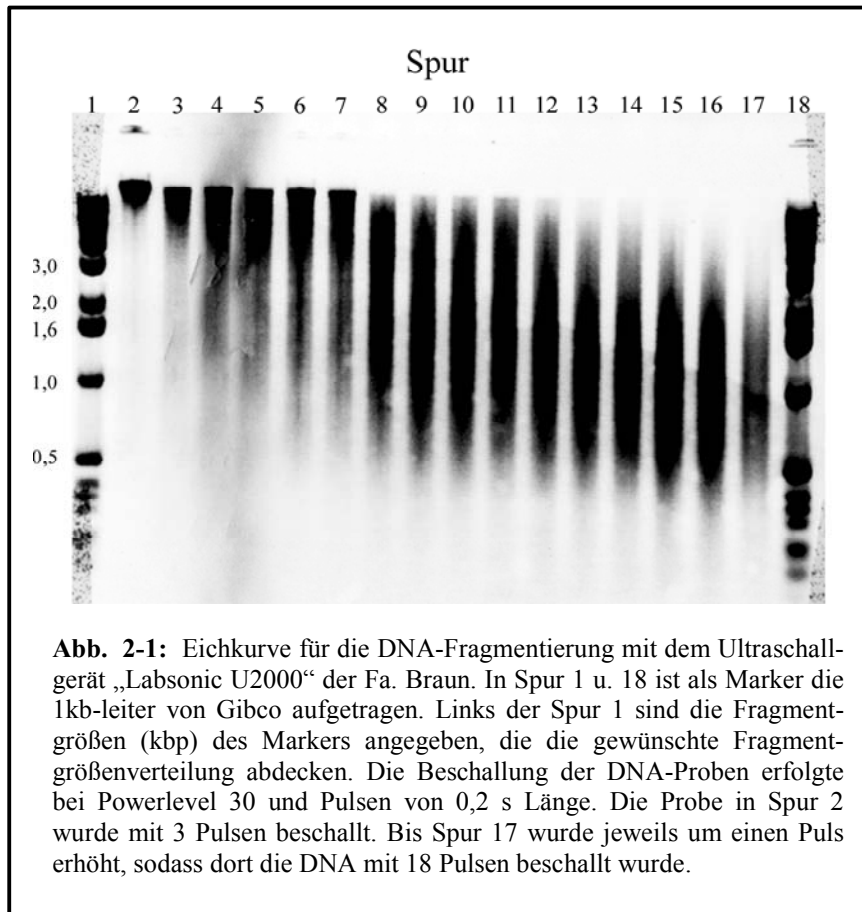
2.1.2 Untersuchung verschiedener Fragmentierungsverfahren

Die Fragmentierungsmethode muss eine statistische Verteilung der Sequenzbruchstücke gewährleisten. Dies ist notwendig um, nach der Sequenzierung einer Anzahl von Bruchstücken, die etwa einer 7-10fachen Abdeckung der ursprünglichen DNA entspricht, die möglichst vollständige Consensussequenz zu erhalten.

Im Rahmen dieser Arbeit wurde zur Fragmentierung der DNA zunächst ein Ultraschallgerät (Kapitel 2.1.2.1) eingesetzt. Auch die Verwendbarkeit eines hydrodynamischen Schersystems (Kapitel 2.1.2.2) wurde getestet.

2.1.2.1 Ultraschallfragmentierung

Für die Fragmentierung der DNA wurde ein Ultraschallgerät der Fa. Braun Modell Labsonic U 2000 mit einer Stabsonde eingesetzt. Erste Fragmentierungsversuche wurden mit der kleinsten einstellbaren Leistungsstufe (Powerlevel 30) kontinuierlich mit verschiedenen Zeitspannen durchgeführt. Mit diesen Parametern führte die Ultraschallbehandlung einer DNA-Probe schon nach 2 Sekunden zu Fragmentgrößen, die unterhalb von 1kbp lagen. Um eine feinere Abstufung in der Fragmentgrößenverteilung zu erreichen, wurde die DNA-Lösung Pulsen von 0,2 Sekunden ausgesetzt. Diese Parameter wurden zur Erstellung einer Eichkurve eingesetzt. Eine Probe von 6µg DNA in 100µL TE-Puffer wurde mit 3 bis 18 Pulsen behandelt. Nach jedem Puls wurde eine gleiche Probenmenge entnommen. Diese Proben wurden dann auf ein Agarosegel aufgetragen und elektrophoretisch (Kapitel 4.9.1) aufgetrennt (Abb. 2-1).



Wie aus der Abbildung zu erkennen konnten mit 11 Ultraschall-Pulsen von 0,2 Sekunden eine Fragmentgrößenverteilung erzielt werden, die den benötigten Bereich von 0,8–3 kbp abdeckt. Die ermittelten Parameter wurden in der Folge für die Herstellung von "shotgun"-Banken eingesetzt. Die Probenvorbereitung und die Durchführung der Fragmentierung mit Ultraschall ist in Kapitel 4.6.3.1 beschrieben. Pro Fragmentierung wurden bis zu 50µg DNA eingesetzt. Im praktischen Einsatz zeigte sich, dass die aus der Eichkurve ermittelten Bedingungen nicht immer zum gewünschten Ergebnis führten. Häufig wurden Fragmentgrößen erhalten, die unterhalb von 1 kbp lagen. In diesen Fällen musste die Fragmentierung wiederholt werden. Die Ultraschallfragmentierung wurde bei insgesamt 12 Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Die Ergebnisse aus dieser Versuchsreihe sind in der Tabelle 2-6 aufgeführt.

Tab. 2-6: Ergebnisse aus den DNA-Fragmentierungen mit dem Ultraschallgerät „Labsonic U2000“ der Fa. Braun.

Klonbezeichnung	Anzahl der Fragmentierungen um die gewünschte Fragmentgrößenverteilung zu erzielen
F18108	1
G1150	2
190A9	1
111K10	1
73M5	2
135E14	2
265A22	4
75G13	3
M17648	5
E24775	3
2A11	3*
11L15	2*

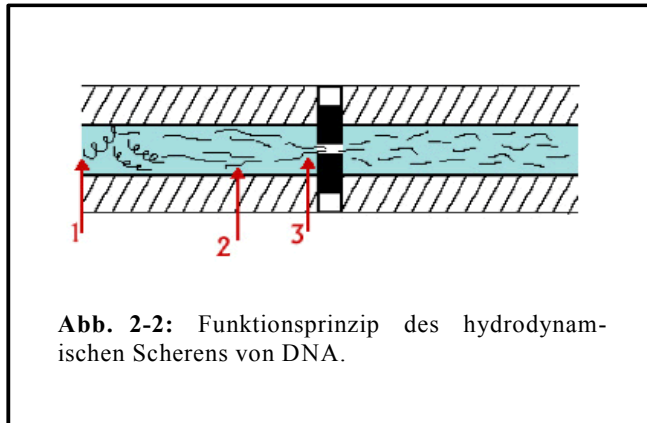
* Bei diesen Klonen erfolgte die Wiederholung der Fragmentierung der DNA durch hydrodynamisches Scheren (Kapitel 2.1.2.2)

Bei nur drei Klonen führte die Fragmentierung mit Ultraschall beim ersten Versuch zu einer Fragmentlängenverteilung, die den benötigten Bereich von etwa 0,8- 3kb abdeckte. Bei allen anderen Klonen musste die Fragmentierung der DNA mindestens einmal wiederholt werden. Insgesamt schlug die Fragmentierung bei den 12 bearbeiteten Klonen 17 mal fehl. Es waren also durchschnittlich pro Klon ca. 2,5 Versuche notwendig, um die gewünschte Fragmentlängenverteilung zu erhalten.

Um diese Fehlerrate reduzieren zu können, wurde ein alternatives Verfahren untersucht, bei dem die Fragmentierung der DNA durch hydrodynamisches Scheren erfolgt (Kapitel 2.1.2.2).

2.1.2.2 Fragmentieren durch hydrodynamisches Scheren

Bei der Fragmentierung von DNA durch hydrodynamisches Scheren handelt es sich um ein neues Verfahren (Oefner P. J. *et al.*, 1996; Hengen P. N., 1997). Die Funktionsweise dieses



Geräts ist relativ einfach (Abb. 2-2). Die gelöste DNA wird durch einen dünnen Schlauch geführt (1). Um den hydrodynamischen Effekt zu erzielen ist der Durchtrittsquerschnitt in dem Schlauch durch eine Säule mit einer hochpräzisen Verengung vermindert. Unmittelbar vor der Präzisionsverengung steigt die Strömungsgeschwindigkeit langsam an (2)

und die DNA wird gedehnt. Beim Eintritt in die Präzisionsverengung nimmt die Strömungsgeschwindigkeit schlagartig stark zu (3). Dadurch wird die DNA bis zum Zerreißen gedehnt und auf diese Weise fragmentiert. Dieser Prozess wird in aufeinander folgenden Zyklen

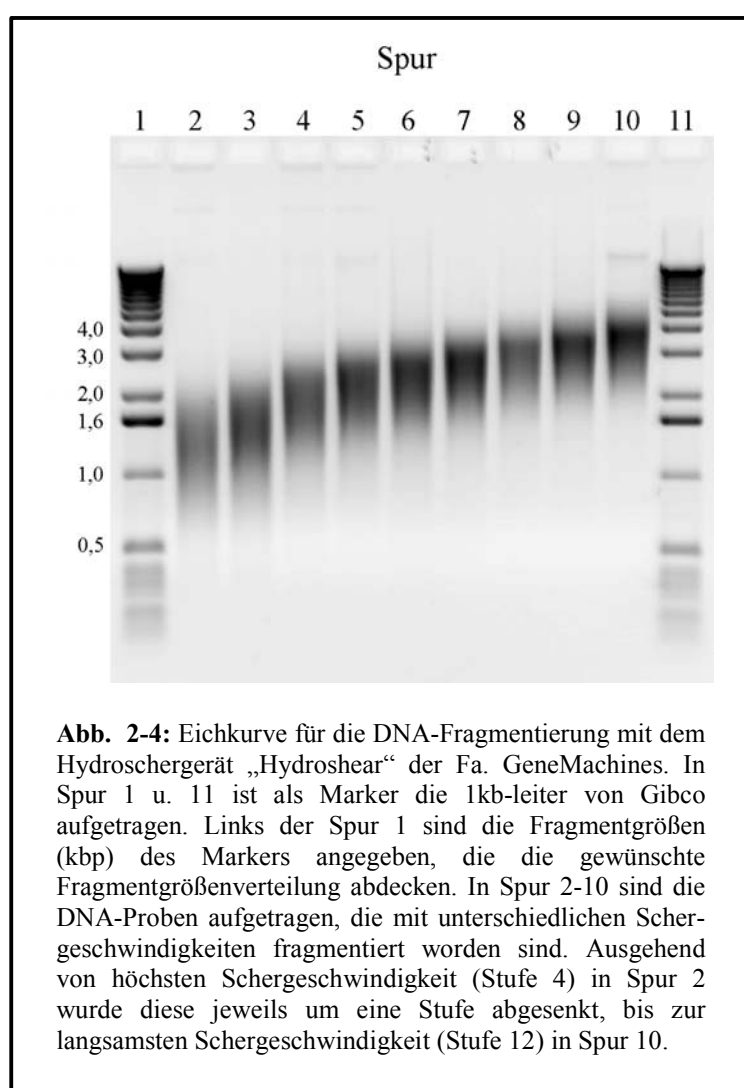


Abb. 2-3: Gezeigt ist das Gerät „HydroShear“ der Fa. GeneMachines.

wiederholt, bis die entstandenen DNA-Bruchstücke so klein sind, dass die Scherkräfte nicht mehr ausreichend groß sind, um die kovalenten Bindungen zwischen den Basen zu brechen. Die Strömungsgeschwindigkeit der DNA-Lösung und der Durchmesser der Verengung legen die abschließenden Fragmentgrößen fest. Auch wiederholte Durchläufe bewirken bei gleicher Durchtrittsöffnung und Strömungsgeschwindigkeit keine weitere Fragmentierung der DNA. Für die Fragmentierung wurde das Gerät der Fa. Gene-Machines Modell „HydroShear“ (Abb. 2-3) eingesetzt.

In ersten Fragmentierungsversuchen führten Schwebstoffe in der DNA-Lösung zum Verstopfen der Präzisionsverengung in der Säule. Die Säule konnte nicht regeneriert werden

und war somit unbrauchbar. Nachfolgend wurde deshalb zur Entfernung von Schwebstoffen aus der DNA-Lösung 10 min bei 14000 UpM zentrifugiert und anschließend der Überstand in ein neues 1,5 ml Eppendorfgefäß umpipettiert. Ein Verstopfen der Säule konnte dadurch sicher vermieden werden. Zur Ermittlung der Geschwindigkeitsstufen, die zu den benötigten Fragmentgrößen führen, wurde eine Eichkurve erstellt. Zu diesem Zweck wurde DNA mit einer Größe von etwa 120 kbp verwendet. Es wurden pro Geschwindigkeitsstufe 3 µg DNA in 100 µl H₂O eingesetzt. Zur Kontrolle der resultierenden Fragmentgröße wurden die Proben auf ein Agarosegel aufgetragen und elektrophoretisch (Kapitel 4.9.1) aufgetrennt. Das Ergebnis ist in der Abb. 2-4 wiedergegeben.



In der Folge wurde zur Herstellung von "shotgun"-Banken für die etwa 1 kbp großen Fragmente am Gerät die Geschwindigkeitsstufe 4 und für die 2,5 – 3 kbp großen Fragmente die Geschwindigkeitsstufe 11 verwendet. Nachfolgend wurde die Fragmentierung von DNA wie in Kapitel 4.6.3.2 durchgeführt. Zur Herstellung von „shotgun“-Banken wurden 8-15 µg DNA

in die Fragmentierung eingesetzt. Die resultierenden Fragmentlängen waren sehr gut reproduzierbar.

Dieses Verfahren wurde zur Fragmentierung der DNA von insgesamt 19 Klonen von Chromosom 21 eingesetzt. Die Ergebnisse aus dieser Versuchsreihe sind in der Tabelle 2-7 aufgeführt.

Tab. 2-7: Ergebnisse der DNA-Fragmentierung unter Verwendung des Geräts „Hydroshear“ der Fa. GeneMachines.

Klonbezeichnung	Anzahl der Fragmentierungen um die gewünschte Fragmentgrößenverteilung zu erzielen
2A11	1
11L15	1
3D19	1
9B17	1
16M16	1
B7L1C4	1
B15L0C0	1
B1L1C6	1
B11L7C6	1
B6L1C6	1
H2137	1
16C2	1
29H4	1
214G14	1
256M13	1
82L10	1
98L15	1
B62L20	1
49E3	1

Wie aus der Tabelle 2-7 ersichtlich, führte dieses DNA-Fragmentierungsverfahren bei jedem der aufgeführten Klone zu der benötigten Fragmentlängenverteilung. Die bei der DNA-Fragmentierung mit Ultraschall (Kapitel 2.1.2.1) vorhandene Fehlerquelle konnte somit erfolgreich beseitigt werden.

2.1.3 Herstellen von DNA-Fragmenten mit glatten Enden und anschließende Fertigstellung von „shotgun“-Banken

Beim Fragmentieren von DNA durch mechanische Einwirkung entstehen sowohl DNA-Bruchstücke mit glatten Enden, als auch DNA-Bruchstücke, deren Enden einzelsträngig-

ge Überhänge aufweisen. Durch enzymatisches Auffüllen in 5'-3' Richtung bzw. Abbauen in 3'-5' Richtung wurden einzelsträngige Überhänge an den Enden der DNA-Fragmente beseitigt. Dadurch wurde gewährleistet, dass eine möglichst große Anzahl an Fragmenten mit glatten Enden in den *Sma* I-verdauten und dephosphorylierten Sequenzierungsvektor pTZ18R (bzw. in den Vektor pUC18) ligiert werden konnte. Hierfür wurde T4-DNA-Polymerase (Denninger *et al.*; 1983) eingesetzt (Kapitel 4.6.4.1), die auch eine 3'-5'-Exonukleaseaktivität besitzt. Anschließend folgte die Ligation (Kapitel 4.6.6) und Transformation (Kapitel 4.4.4). Dieses Verfahren wurde bei 16 Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Konnte keine ausreichende Anzahl von transformierten Bakterienkolonien erhalten werden, um eine "shotgun"-Bank zu komplettieren, wurde der Versuch wiederholt. Die Ergebnisse aus dieser Versuchsreihe sind in der Tabelle 2-8 aufgeführt.

Tab. 2-8: Ergebnisse zur Auffüllreaktion mit T4-DNA-Polymerase. Neben der Klonbezeichnung ist die Anzahl der Versuche angegeben, die zur Komplettierung einer „shotgun“-Bank notwendig waren.

Klonbezeichnung	Anzahl der Versuche
F18108	2
G1150	1
190A9	1
111K10	1
73M5	1
135E14	1
265A22	1
75G13	2
M17648	1
E24775	1
2A11	4*
11L15	3*
3D19	3*
9B17	1*
16M16	4*
B7L1C4	2*

*Bei diesen Klonen erfolgte die DNA-Fragmentierung durch hydrodynamisches Scheren (Kapitel 2.1.2.2).

Wie aus der Tabelle 2-8 ersichtlich, erwies sich dieses Verfahren anfänglich als relativ unproblematisch. Ab dem Klon 2A11 ist allerdings eine Häufung der notwendigen Versuchswiederholungen zu erkennen. Diese erhöhte Anzahl von Fehlversuchen trat mit dem Wechsel des DNA-Fragmentierungsverfahrens auf. Anstelle der Fragmentierung durch Ultraschall (Kapitel 2.1.2.1) erfolgte diese ab dem Klon 2A11 durch hydrodynamisches Scheren

(Kapitel 2.1.2.2). Bei diesem Protokoll (Kapitel 4.6.3.2) wurde die zur Fragmentierung eingesetzte DNA-Menge auf ein Drittel reduziert. Offensichtlich führten größere DNA-Mengen (Kapitel 4.6.3.1) schon bei der Fragmentierung zu einer ausreichenden Anzahl an DNA-Bruchstücken, um nachfolgend eine komplette „shotgun“-Bank zu erhalten. Dadurch fiel die schlechte Reaktionsausbeute an Fragmenten mit glatten Enden, die durch die enzymatische Reaktion mit T4-DNA-Polymerase erzeugt wurden, nicht ins Gewicht.

Deshalb wurde für den Schritt der enzymatischen Herstellung glatter Fragmentenden ein alternatives Protokoll (Kapitel 4.6.4.2) auf seine Eignung untersucht. Bei diesem Protokoll wird neben der T4-DNA-Polymerase noch zusätzlich Klenow-Fragment in den Reaktionsansatz eingesetzt. Das Klenow-Fragment besitzt, wie die T4-DNA-Polymerase, sowohl 5'-3'-Polymeraseaktivität als auch 3'-5'-Exonukleaseaktivität. Im Anschluss an dieses Protokoll folgte die Ligation (Kapitel 4.6.6) und die Transformation (Kapitel 4.4.4).

Dieses Verfahren wurde bei 17 Klonen, die ein Fragment des humanen Chromosoms 21 trugen, eingesetzt. Die Ergebnisse aus dieser Versuchsreihe sind in der Tabelle 2-9 aufgeführt.

Tab. 2-9: Ergebnisse zur Auffüllreaktion mit T4-DNA-Polymerase und Klenow-Fragment. Neben der Klonbezeichnung ist die Anzahl der Versuche angegeben, die zur Komplettierung einer „shotgun“-Bank notwendig waren.

Klonbezeichnung	Anzahl der Versuche
11L15	1
3D19	1
16M16	1
B7L1C4	1
B15L0C0	1
B1L1C6	1
B11L7C8	1
B6L1C6	1
H2137	1
16C2	1
29H4	1
214G14	1
256M13	1
82L10	1
98L15	1
B62L20	1
49E3	1

Dieses Protokoll erwies sich als sehr zuverlässig. Bei jedem der aufgeführten Klone wurde im ersten Versuch eine ausreichende Anzahl an Transformanden erhalten, um die „shotgun“-Bank zu komplettieren.

In dieser Versuchsreihe wurde zusätzlich zur Steigerung der Anzahl an Transformanden der in die Transformation (Kapitel 4.4.4) eingesetzte Bakterienstamm gewechselt. Anstelle des bisher verwendeten Bakterienstamms XL1Blue wurde DH10B eingesetzt. Dieser Bakterienstamm weist eine 10-15fach höhere Transformationsrate auf. Die Anzahl an erhaltenen Transformanden konnte entsprechend gesteigert werden. Allerdings zeigte sich bei der Kultivierung der DH10B-Zellen auf Agarplatten mit einer IPTG-Konzentration 200 μ M/l und einer X-Gal-Konzentrationen von 40 mg/l eine ausgeprägte Satellitenkoloniebildung. Durch Senkung der IPTG-Konzentration auf 80 μ M/l und Erhöhung der X-Gal-Konzentration auf 70 mg/l konnte die Satellitenkoloniebildung beseitigt werden. Durch die Verwendung des Bakterienstamms DH10B für die Transformation konnte die ausgangs eingesetzte DNA-Menge, die zur erfolgreichen Herstellung von "shotgun"-Banken nötig war, von ca. 15 μ g auf ca. 8 μ g gesenkt werden.

Die vollautomatische Übertragung der Bakterienkolonien von Agarplatten in Mikrotiterplatten im 384er Format erfolgte mit dem Picking- und Spotting-Roboter Modell „Q-Bot“ der Fa. Genetix. Die verwendeten Einstellungsparameter sind in Kapitel 4.7 angegeben.

2.1.4 Zusammenfassung der Optimierungen zur Herstellung von "shotgun"-Banken

Im anfänglich eingesetzten Herstellungsverfahren von „shotgun“-Banken wurde das Fällungsverfahren (Kapitel 2.1.1.1) zur DNA-Präparation eingesetzt. Im Anschluss an die DNA-Fragmentierung durch Ultraschall (Kapitel 2.1.2.1) folgte die Herstellung glatter Fragmentenden mit T4-DNA-Polymerase (Kapitel 2.1.3). Mit diesem Verfahren wurden für die parallele Herstellung von zwei „shotgun“-Banken insgesamt 12 Arbeitstage benötigt.

Durch schrittweise Optimierung der DNA-Präparation (Kapitel 2.1.1.2.1 bis Kapitel 2.1.1.2.3) konnten sechs „shotgun“-Banken innerhalb von 8 Arbeitstagen parallel hergestellt werden. Der Anteil an chromosomaler DNA aus *E. coli* in den „shotgun“-Banken konnte durchschnittlich auf ca. 6% gesenkt werden. Durch unzureichend reproduzierbare Ergebnisse bei den einzelnen Herstellungsschritten waren für die erfolgreiche Herstellung einer „shotgun“-Bank durchschnittlich ca. 4 Versuche notwendig.

Durch Etablierung eines hydrodynamischen Schersystems zur DNA-Fragmentierung (Kapitel 2.1.2.2) und Einführung eines verbesserten Protokolls zur enzymatischen Herstellung glatter Fragmentenden (Kapitel 2.1.3) konnte die Reproduzierbarkeit dieser Teilschritte deut-

lich verbessert werden. Durch diese Verfahren konnte zusätzlich die notwendige DNA-Menge zur erfolgreichen Herstellung von „shotgun“-Banken verringert werden. Dadurch konnte eine weitere Optimierung des DNA-Präparationsverfahrens (Kapitel 2.1.1.2.4) erfolgen. Wie zuvor konnten mit diesem Herstellungsverfahren sechs „shotgun“-Banken parallel hergestellt werden. Der zeitliche Aufwand konnte allerdings auf 6 Arbeitstage verkürzt werden. Durch das verbesserte DNA-Präparationsprotokoll konnte der Anteil an chromosomaler DNA aus *E. coli* in den Banken nochmals reduziert werden. Er beträgt durchschnittlich 3,1%. Ferner konnte die Reproduzierbarkeit des Herstellungsverfahrens gesteigert werden. Für die erfolgreiche Herstellung einer „shotgun“-Bank waren durchschnittlich nur noch 1,4 Versuche notwendig.

In der Summe der erarbeiteten Optimierungen konnte ein sehr zuverlässiges Herstellungsverfahren für „shotgun“-Banken etabliert werden. Dieses Verfahren wurde in dieser Form in den Routinebetrieb übernommen.

2.2 Hybridisierungsexperimente zur Entfernung des Klonierungsvektors aus „shotgun“-Banken

Wie einleitend erwähnt (Kapitel 1.4) gibt es bei der „hierarchical shotgun“-Sequenzierung verfahrensbedingt in einer „shotgun“-Bank Klone, die als Insertion im Sequenzierungsvektor Teile des bakteriellen Genoms bzw. des Klonierungsvektors tragen. Diese werden anschließend mitsequenziert. Der Anteil von Klonen in einer „shotgun“-Bank, die Teile des Vektors als Insertion im Sequenzierungsvektor tragen, ergibt sich aus dem Größenverhältnis von Klonierungsvektor zu Insertion und Vektor. Die Tabelle 2-10 zeigt die Größen der Klonierungsvektoren und ihrer Insertionen, sowie ihr Verhältnis zueinander. Die aufgeführten Daten sind aus den, an der GBF Braunschweig/Abt. Genomanalyse bearbeiteten, „shotgun“-Banken der Chromosomen 9 und 21 des Humangenoms ermittelt worden.

Der Anteil an Klonierungsvektor tragenden Klonen in diesen Banken betrug minimal 4% und maximal 26%. Über alle Banken gemittelt lag der durchschnittliche Anteil bei ca. 9,3%. Die Kontamination mit bakterieller DNA konnte durch Optimierung der DNA-Präparation (Kapitel 2.1) auf einen durchschnittlichen Wert von 3,1% gesenkt werden. Dies bedeutet, dass die Kontamination einer „shotgun“-Bank durch Klonierungsvektor durchschnittlich dreimal so hoch liegt wie die Kontamination durch bakterielle DNA. In Einzelfällen kann dieser Wert mehr als das achtfache betragen. Deshalb wurde eine Methode entwickelt die Klone, die als Insertion Klonierungsvektor-DNA tragen, vor der Sequenzierung durch Hybridisierung mit DIG-markierten Sonden zu identifizieren und anschließend aus der Bank zu entfernen.

Für eine vollständige Entfernung des Klonierungsvektors dürfen keine Sequenzhomologien zum verwendeten Sequenzierungsvektor vorhanden sein. In diesem Fall müssten die sequenzhomologen Bereiche bei der Sondenherstellung ausgeschlossen werden, um falsch positive Signale auszuschließen. Durch Sequenzvergleiche zwischen den jeweiligen Klonierungsvektoren (Tab. 2-10) und den Sequenzierungsvektoren pTZ18R und pUC18R konnten allerdings keine Sequenzhomologien gefunden werden.

Tab. 2-10: Die Tabelle zeigt die Klonierungsvektoren und den minimalen, maximalen und durchschnittlichen Anteil an Klonierungsvektor-DNA tragenden Klonen in den „shotgun“-Banken von Chromosom 9 und 21 des Humangenoms, die an der GBF Braunschweig/Abt. Genomanalyse bearbeitet wurden. Quelle der Klonierungsvektordaten: www.rzpd.de/general/html/glossary/vectors/vector.html.

Vektor (Anzahl der Banken)	Vektor- größe (kbp)	Insertions- größe Min. – Max. (kbp)	Anteil in einer "shotgun"- Bank Min. – Max. (%)	Insertions- größe Durchschnitt (kbp)	Anteil in einer "shotgun"- Bank Durchschnitt (%)
pCYPAC2N (24)	16	45 - 178	8 - 26	129	11
pBeloBAC11 (11)	7,5	30 - 119	6 - 20	80	9
pBACe3.6 (32)	8,7	78 - 212	4 - 10	158	5
Lawrist 4 u. 16 (12)	8,2	27 - 46	15 - 23	37	18

Für die experimentelle Untersuchung des Verfahrens wurden der ca. 191 kbp große Klon 381K22 und der ca. 212 kbp große Klon 512L9 aus dem aktuellen Sequenzierprojekt von Chromosom 9 verwendet. Beide Klone stammen aus der selben Kartierungsbank, für deren Herstellung der 8,7 kbp große Klonierungsvektor pBACe3.6 verwendet wurde. Um den gesamten Klonierungsvektor abzudecken, wurden Primer (Kapitel 4.3.4) für neun 0,8-1 kbp große Sonden geplant. Die Herstellung der Sonden erfolgte mittels PCR (Kapitel 4.8.2), als Markierung diente DIG-11-dUTP. Für die Hybridisierungsversuche wurden „shotgun“-Banken dieser Klone hergestellt, die jeweils 3 Mikrotiterplatten im 384er-Format umfassten. Von diesen Banken wurden jeweils Hybridisierungsfilter hergestellt (Kapitel 4.11.1). Da Sondengemische zu falsch positiven Signalen führen können (mündl. Mitt. Labor „Lion Bioscience“ in Heidelberg) erfolgte die Hybridisierung auf zwei Arten. Zum einen wurde jeweils eine einzelne Sonde, zum anderen wurden alle Sonden als Gemisch in die Hybridisierungsreaktion (Kapitel 4.11.2) eingesetzt. Abbildung 2-5 zeigt das Hybridisierungsergebnis für den Klon 512L9. Für die Hybridisierung wurde das Sondengemisch eingesetzt.

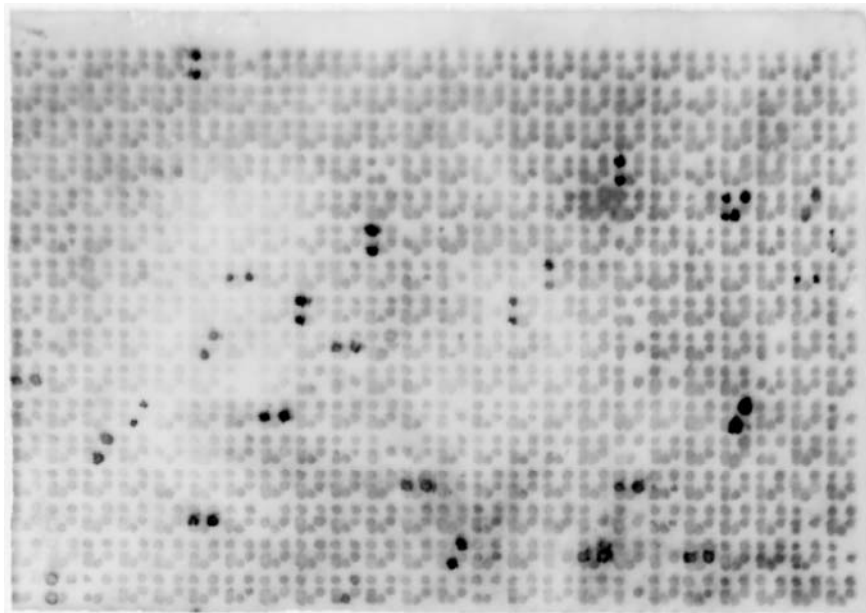
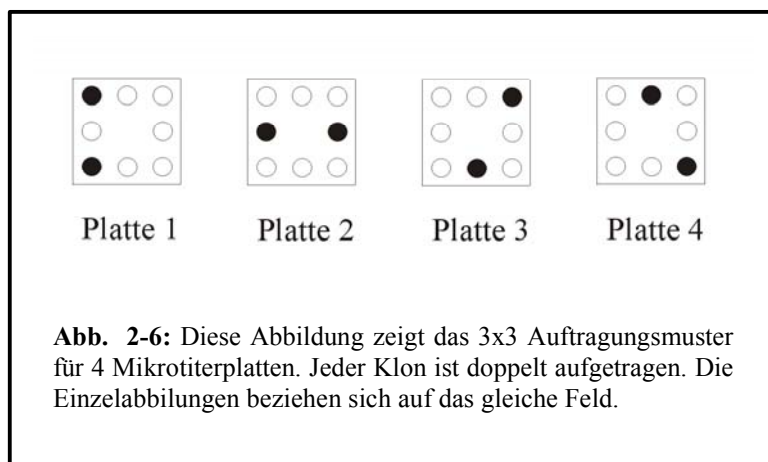


Abb. 2-5: Gezeigt ist das Hybridisierungsergebnis der Identifizierung Klonierungsvektor-DNA tragender Bakterienkolonien in der „shotgun“-Bank des Klons 512L9. In die Hybridisierung wurde ein Sondengemisch eingesetzt

Jeder Klon einer Mikrotiterplatte wurde zweimal nach einem vorgegebenen Muster auf den Filter übertragen (Abb. 2-6). Durch dieses Muster kann festgestellt werden von welcher Mikrotiterplatte der identifizierte Klon stammt.



Sowohl die Hybridisierung mit Einzelsonden als auch mit dem Sondengemisch führten zu dem gleichen Ergebnis, 30 Klone der Bank 381K22 und 28 Klone der Bank 512L9 zeigten ein positives Signal. In der Tabelle 2-11 ist die Anzahl positiver Signale für jede der 3 Mikro-

titerplatten einzeln aufgeführt. Bezogen auf die jeweils 1152 Klone dieser Banken entspricht dies einem Anteil von ca. 2,6% in der Bank 381K22 und 2,4% in der Bank 512L9. Errechnet wurde für den Klon 381K22 ein Anteil von 4,1% (ca. 47 Klone) und für den Klon 512L9 ein Anteil von 4,5% (ca. 52 Klone).

Tab. 2-11: In der Tabelle sind die Positionen der Klone der „shotgun“-Banken angegeben, die mittels der Hybridisierungsexperimente als Klonierungsvektor tragende Klone identifiziert worden sind. Ferner sind die Positionen von nicht angewachsenen Klonen angegeben.

	Anzahl positiver Signale		
Klonname	Mikrotiterplatte 1	Mikrotiterplatte 2	Mikrotiterplatte 3
381K22	9	9	12
512L9	9	11	8

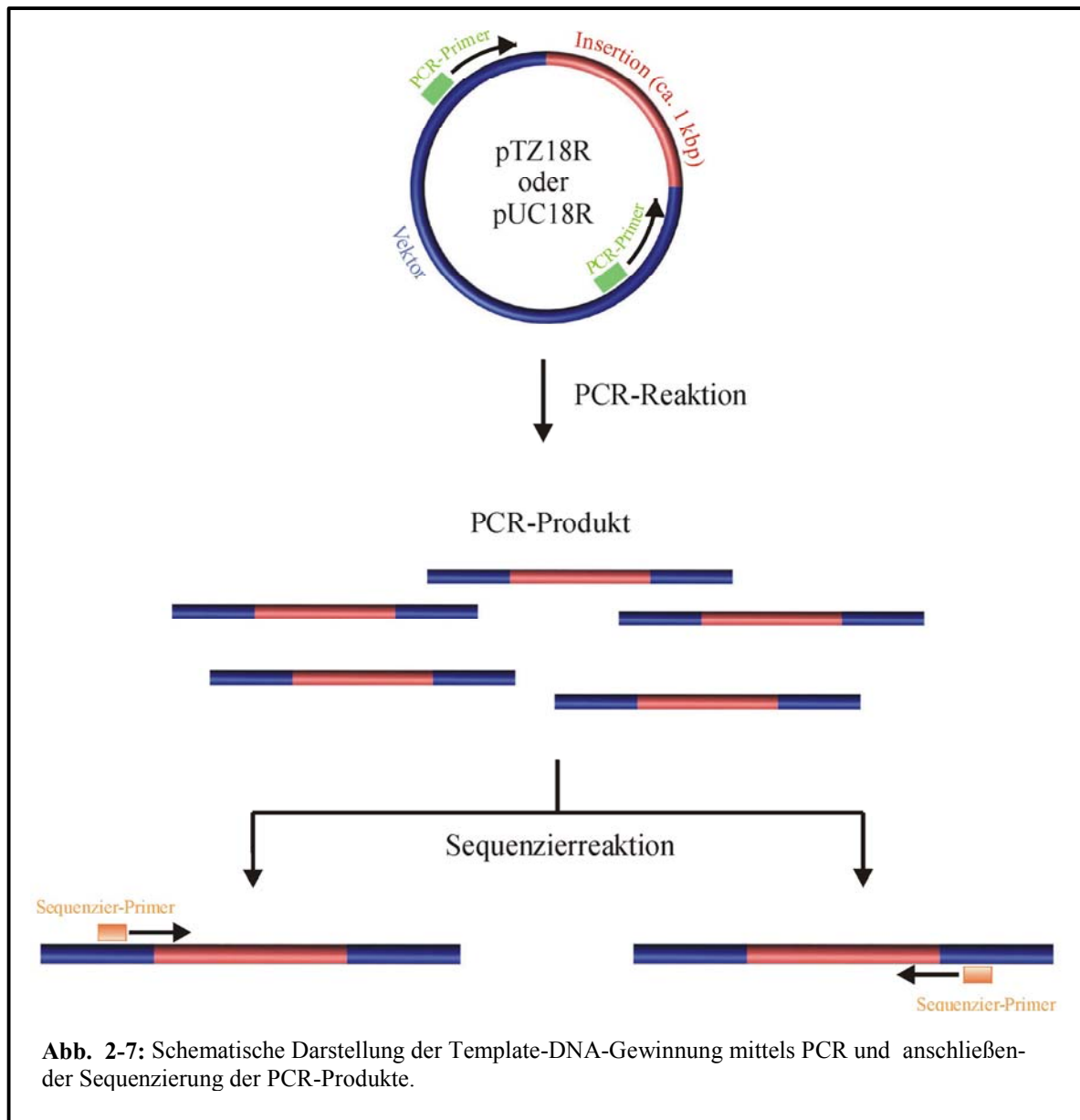
Bei den Mikrotiterplatten 1 beider Banken wurden vor der Sequenzierung die durch Hybridisierung identifizierten Klone durch einen Umorganisationsschritt entfernt und durch Klone der Mikrotiterplatten 3 ersetzt. In gleicher Weise wurde mit nicht angewachsenen Klonen der Mikrotiterplatten 1 verfahren. Die Mikrotiterplatten 2 wurden unverändert beibehalten. Die Mikrotiterplatten 1 und 2 der beiden Banken wurden anschließend in die Sequenzierung eingesetzt.

Die Auswertung der Sequenzierungsdaten der Mikrotiterplatten 1 und 2 zeigte, dass es keine Klone gab, die Teile des Klonierungsvektors trugen und nicht bei der Hybridisierungsreaktion erkannt wurden. Bei den Mikrotiterplatten 2 war es zusätzlich möglich Klone zu identifizieren, die fälschlicherweise bei der Hybridisierung ein Signal gaben, obwohl sie keinen Klonierungsvektor trugen. Nur im Fall des Klons 512L9 wurden 3 Klone als falsch positiv identifiziert.

Aus den „shotgun“-Banken der Klone 381K22 und 512L9 konnten somit Klonierungsvektor-DNA tragende Bakterienklone vollständig durch Hybridisierung entfernt werden.

2.3 DNA-Matrizengewinnung aus Bakterienzellen mittels PCR zum Einsatz in die Sequenzierungsreaktion

Handelsübliche Plasmid-DNA-Präparationskits besitzen beim Einsatz in der Megasequenzierung zwei entscheidende Nachteile. Einerseits verursachen sie Kosten, die einen erheblichen Anteil an den Gesamtaufwendungen der Sequenzierung ausmachen. Andererseits können nur 96 und nicht 384 Proben parallel prozessiert werden. Diese beiden Nachteile lassen sich mittels der Polymerase-Kettenreaktion beseitigen (Abb. 2-7). Versuche haben gezeigt, dass



unter Standard-Bedingungen (Kapitel 4.8.1) auch dann ein spezifisches PCR-Produkt erhalten wird, wenn anstelle von isolierter DNA Bakterienzellen im PCR-Ansatz eingesetzt werden. Für diese Versuche wurden Klone von „shotgun“-Banken verwendet, für deren Herstellung der Vektor pTZ18R eingesetzt wurde. Die PCR-Primer binden ca. 200 bp vor der Insertion. Das PCR-Produkt konnte mit einer durchschnittlichen Ausbeute von 750 ng aus einer Polymerase-Kettenreaktion mit 20 µl Reaktionsansatz gewonnen werden. Das entspricht einer Konzentration von 37,5 ng/µl.

Vor den ersten Sequenzierungsversuchen wurde die einzusetzende PCR-Produkt-Menge berechnet. Das PCR-Produkt ist mit ca. 1,4 kbp um den Faktor 2,7 kleiner als die Sequenzierungsvektoren pTZ18R und pUC18R inklusive Insertion. Ausgehend von den 600-800 ng Plasmid-

DNA, die in die Sequenzierreaktion (Kapitel 4.8.3) eingesetzt werden ergibt sich für das PCR-Produkt eine einzusetzende Menge von 220-300 ng.

Erste Sequenzierungsversuche, bei denen 220 ng des PCR-Produkts direkt in die Sequenzierung eingesetzt wurden, zeigten jedoch kurze Leseweiten von ca. 100 bp und überlagerte Signale. Die Signalintensität im Elektropherogramm war bei diesen Versuchen verhältnismäßig schwach. Dies machte es unmöglich, die Signale eindeutig einer durchgehenden Sequenz zuzuordnen. Diese Sequenziererergebnisse werden wahrscheinlich durch die Primer und die Desoxynukleotide aus dem PCR-Ansatz verursacht, die mit dem PCR-Produkt in die Sequenzierreaktion eingebracht werden. Die aus der PCR stammenden Primer führen bei der Sequenzierreaktion zu zusätzlichen Startpunkten für die DNA-Polymerase. Dies kann zu überlagerten Signalen führen. Die verringerte Signalintensität wird wahrscheinlich durch die zusätzlichen Desoxynukleotide verursacht, die das Verhältnis von fluoreszenzmarkierten Abbruchmolekülen (ddNTP's) zu dNTP's im Sequenzieransatz verringern. Bei der Sequenzierreaktion wird deshalb seltener ein fluoreszenzmarkiertes Abbruchmolekül eingebaut. Die Folge ist eine verringerte Signalintensität bei der Detektion durch den Sequenzierautomaten.

Für die Entfernung der PCR-Primer und der Desoxynukleotide wurden drei Möglichkeiten in Betracht gezogen.

1.) Kommerzielle Filterprodukte:

Hierbei werden durch Größenselektion Primer und Nukleotide vom PCR-Produkt abgetrennt.

2.) Einsatz von Biotin endmarkierten Primern:

In die PCR-Reaktion werden Biotin endmarkierte Primer eingesetzt. Im Anschluss wird das PCR-Produkt über das Biotin an die Streptavidinoberfläche ferromagnetischer Kügelchen gebunden („Dynabeads“). Im magnetischen Feld kann so das PCR-Produkt von der Lösung abgetrennt und aufgereinigt werden.

3.) DNA-Fällung mit Alkohol:

Durch Fällung des PCR-Produkts mit Ethanol bzw. Isopropanol werden die PCR-Primer und Desoxynukleotide abgetrennt.

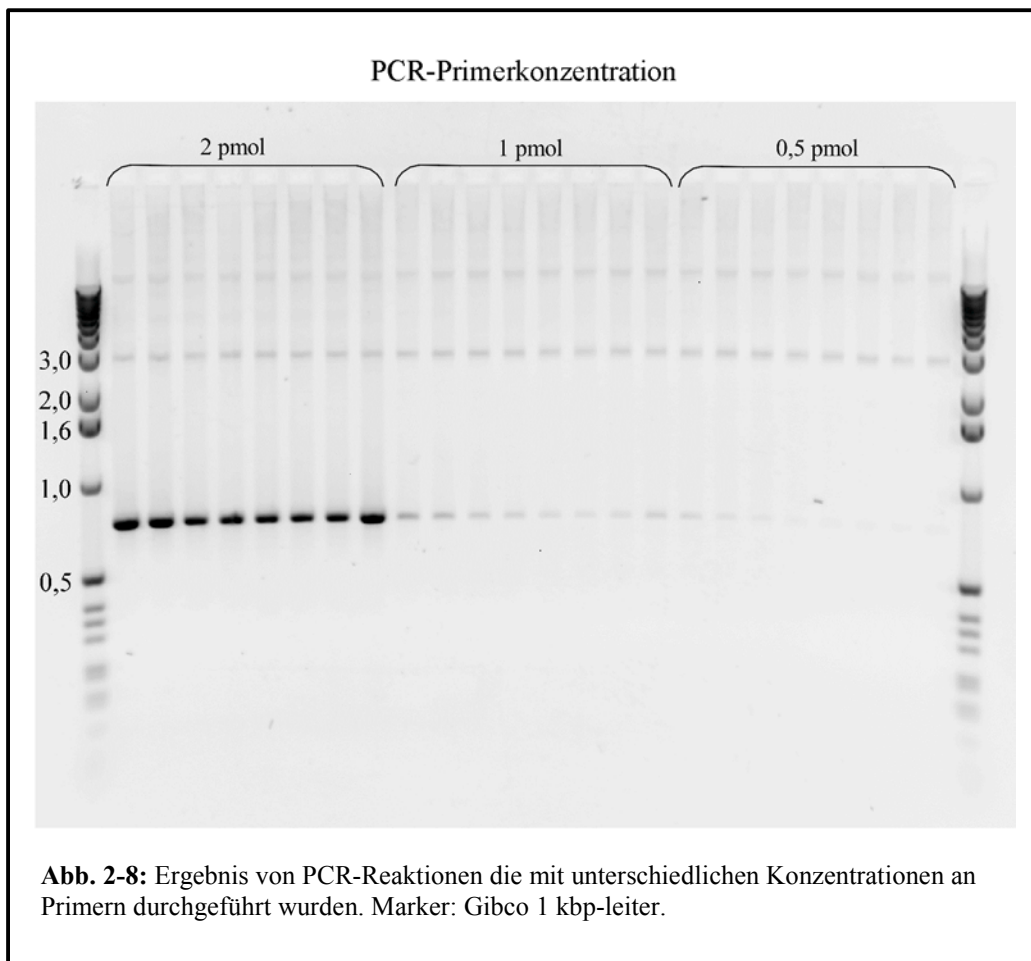
Im Vergleich zu DNA-Plasmidpräparations-Kits bieten kommerzielle Filterprodukte keine nennenswerten Vorteile. Sie sind ebenfalls auf ein 96er-Raster beschränkt und erfordern einen ähnlichen Kosten- und Arbeitsaufwand. Die Verwendung von Biotin endmarkierten PCR-Primern in Kombination mit „Dynabeads“ hat den Vorteil, dass 384 Proben parallel pro-

zessiert werden können und sich der Arbeitsaufwand deutlich verringert. Allerdings sind „Dynabeads“ teuer und verursachen deshalb einen zusätzlichen Anteil an den Gesamtaufwendungen der Sequenzierung. Die Aufreinigung der PCR-Produkte durch Fällung mit Alkohol bietet für die nachfolgenden Untersuchungen die besten Voraussetzungen. Diese Methode erlaubt die parallele Prozessierung von 384 Proben und verursacht geringe Kosten.

Die Aufreinigung des PCR-Produkts mittels Alkoholfällung muss zwei Kriterien erfüllen:

- 1.) Die PCR-Primer müssen so weit abgetrennt werden, dass die verbleibende Restmenge keinen störenden Einfluss auf die nachfolgende Sequenzierreaktion hat.
- 2.) Das aufgereinigte PCR-Produkt muss in ausreichender Menge erhalten werden, um eine Sequenzierung durchführen zu können.

Um abzuschätzen, ab welcher Konzentration der Einfluss der PCR-Primer auf die Sequenzierreaktion vernachlässigbar ist wurden PCR-Reaktionen (Kapitel 4.8.1) mit verschiedenen Primer-Konzentrationen (2 pmol, 1 pmol und 0,5 pmol je Primer) durchgeführt. Wie aus der Ab-



bildung 2-8 zu erkennen, erhält man ab einer Konzentration von 1 pmol je PCR-Primer im Vergleich zu 2 pmol je PCR-Primer nur noch eine sehr geringe Menge an Produkt. Überträgt man dieses Ergebnis auf die Sequenzierreaktion, bei der 5 pmol je Sequenzier-Primer eingesetzt werden, so kann man davon ausgehen, dass ab einer Konzentration von 1 pmol je PCR-Primer der störende Einfluss vernachlässigbar ist. Da in eine PCR-Reaktion 5 pmol je Primer eingesetzt werden, müssen bei der Fällung mindestens 80% der Primer abgetrennt werden. Für nachfolgende Fällungsversuche (Kapitel 4.4.6.1) mit Ethanol bzw. Isopropanol wurden Primer mit einer Konzentration von 0,5 pmol / μ l eingesetzt. Diese Primermenge entspricht 5 pmol je Primer in einem PCR-Ansatz mit einem Reaktionsvolumen von 20 μ l. Nach einer einheitlichen Inkubationsdauer folgte eine Zentrifugation (2550 x g) für 20 min, 30min bzw. 40min. Tabelle 2-12 zeigt die Ergebnisse zu diesen Versuchen.

Tabelle 2-12: Ergebnisse aus den Fällungsversuchen zur Abtrennung von PCR-Primern.

Probe	OD _{260nm} Probe 1	OD _{260nm} Probe 2	OD _{260nm} Ø	Primerrestmenge %
Standard	0,080	0,076	0,0780	100
Isopropanol 20 min	0,013	0,015	0,0140	17,9
Isopropanol 30 min	0,018	0,016	0,0170	21,8
Isopropanol 40 min	0,015	0,016	0,0155	19,9
Ethanol 20 min	0,027	0,032	0,0295	37,8
Ethanol 30 min	0,028	0,029	0,0285	36,5
Ethanol 40 min	0,030	0,032	0,0310	39,7

Durch Fällung mit Isopropanol konnten ca. 80% der ursprünglich eingesetzten Primer abgetrennt werden. Bei der Fällung mit Ethanol konnte nur ca. 62% des Primers entfernt werden. Die Zentrifugationsdauer hatte in dem gewählten Zeitintervall keine Auswirkungen auf das Ergebnis.

Als nächstes wurde untersucht, in welchen Ausbeuten sich das PCR-Produkt unter den zuvor angewendeten Fällungsbedingungen wiedergewinnen lässt. Die Gewinnung einer ausreichenden Menge aufgereinigten PCR-Produkts ist sehr aufwendig. Deshalb wurde anstelle des PCR-Produkts linearisierte Vektor-DNA (pTZ18R, *Sma*I-verdaut) mit der zuvor errechneten Konzentration von 37,5 ng/ μ l in die Fällungsreaktion (Kapitel 4.4.6.1) eingesetzt. Tabelle 2-13 zeigt die Ergebnisse dieser Versuche.

Ergebnisse

Tabelle 2-13: Ergebnisse der Versuche zur Ermittlung der erzielbaren Restmenge an DNA des Vektors pTZ18R, die sich durch Fällung bei unterschiedlichen Bedingungen zurückgewinnen lässt.

Probe	OD _{260nm} Probe 1	OD _{260nm} Probe 2	OD _{260nm} Ø	Restmenge des Vektors pTZ18R (%)
Standard	0,468	0,463	0,4655	100
Isopropanol 20 min	0,268	0,273	0,2705	58,1
Isopropanol 30 min	0,273	0,284	0,2785	59,8
Isopropanol 40 min	0,276	0,266	0,2710	58,2
Ethanol 20 min	0,255	0,259	0,2570	55,2
Ethanol 30 min	0,275	0,280	0,2775	59,6
Ethanol 40 min	0,263	0,277	0,2700	58,0

Sowohl mit Isopropanol als auch mit Ethanol konnte das PCR-Produkt mit einer Ausbeute von ca. 60% nach der Fällung erhalten werden. Bezogen auf eine Ausgangsmenge von 750 ng werden also 450 ng PCR-Produkt zurückgewonnen. Die Zentrifugationsdauer hatte im gewählten Zeitintervall keine Auswirkung auf das Ergebnis.

Die Ergebnisse der einzelnen Versuche zeigen, dass die Fällungsmethode mit Isopropanol und anschließender Zentrifugation für 20 min gut für die Aufreinigung des PCR-Produkts geeignet ist. Das PCR-Produkt wird mit Ausbeuten von ca. 450 ng zurückgewonnen. Diese Menge ist ausreichend um einen Klon von beiden Seiten sequenzieren zu können. Dadurch halbiert sich die Restmenge an PCR-Primern. Somit werden nur noch 0,5 pmol je PCR-Primer mit dem PCR-Produkt in die Sequenzierreaktion eingebracht. Wie in Abbildung 2-8 gezeigt ist von dieser Menge an PCR-Primern kein störender Einfluss auf das Sequenziererergebnis mehr zu erwarten.

Nach diesen Vorversuchen wurde von unterschiedlichen Klonen mittels Polymerase-Kettenreaktion PCR-Produkt gewonnen und wie beschrieben durch Fällung mit Isopropanol aufgereinigt. Von jedem der aufgereinigten PCR-Produkte wurde ein Aliquot auf ein Agarosegel aufgetragen, elektrophoretisch aufgetrennt und nachfolgend auf Nebenprodukte untersucht. Bei keinem der Klone kam es während der PCR zur Bildung von Nebenprodukten. Das restliche PCR-Produkt wurde für die Sequenzierung des Plus- bzw. des Minusstrangs eingesetzt. In die Sequenzierreaktion (Kapitel 4.8.3) wurden unterschiedliche Mengen an PCR-Produkt (90-180 ng) eingesetzt. Zusätzlich wurde die Anzahl der Zyklen (45 bzw. 37) bei der Sequenzierreaktion variiert. Optimale Ergebnisse ließen sich mit 90-120 ng PCR-Produkt und 37 Zyklen bei der Sequenzierreaktion erzielen. Einige Beispiele der so erhaltenen Sequenzierungsergebnisse sind im Anhang A (Kapitel 7.1) aufgeführt.

Mit diesem Verfahren wurden Sequenzierergebnisse erzielt, die qualitativ gleichwertig zur Standardmethode sind, bei der Plasmid-DNA in die Sequenzierreaktion eingesetzt wird. Somit ist es gelungen, eine kostengünstige Methode zu etablieren, mit der die Nachteile handelsüblicher Plasmid-DNA-Präparationskits in der Megasequenzierung beseitigt werden konnten.

2.4 Sequenzierung und Annotierung von Klonen des Chromosoms 21

Im Rahmen dieser Arbeit wurden die benachbarten Klone 98L15 und 49E3 von Chromosom 21 bearbeitet. Diese Klone sind im Übergangsbereich der Arme 21q11.1 und 21q11.2 des Chromosoms lokalisiert. Im Anschluss der vollständigen Sequenzierung dieser Klone folgte die Annotierung, um so mögliche Gene oder bekannte Gene aufzufinden.

Im ersten Schritt der Sequenzierung wurde eine ausreichende Anzahl von Rohsequenzen für eine 7-10-fache Abdeckung des Klons 98L15 bzw. 49E3 erzeugt. Die Sequenzierung von Klonen mit einer 1 kbp großen Insertion erfolgte mit dem „ABITM377“ (Applied Biosystems) und die der Klone mit 2-2,5 kbp großer Insertion mit dem „dna sequencer LONG READIR 4200“ (LI-COR). Die Durchführung der jeweiligen Sequenzierreaktionen erfolgte nach den Protokollen in Kapitel 4.8.3. Die Vorbereitung der Sequenziergele ist in Kapitel 4.9.2, und die Bedingungen für den folgenden Sequenzierlauf sind in Kapitel 4.9.2.1 beschrieben. Von der „shotgun“-Bank des Klons 98L15 (Kapitel 2.1.1.2.4) wurden 1152 Klone mit einer ca. 1 kbp großen Insertion und 192 Klone mit einer 2-2,5 kbp großen Insertion jeweils von beiden Seiten sequenziert. Die von einer Seite sequenzierten 96 Testklone (Kapitel 2.1.1.2.4) wurden diesen Rohsequenzen zugefügt. Von der „shotgun“-Bank des Klons 49E3 (Kapitel 2.1.12.4) wurden 1824 1 kbp- und 288 2-2,5 kbp-Klone sequenziert. Zusätzlich lagen Rohsequenzen von 768 Testklonen (Kapitel 2.1.1.2.4) vor. Die so erhaltenen Rohsequenzen dieser beiden Klone wurden wie in Kapitel 4.10 beschrieben weiter bearbeitet. Zuerst wurden die Sequenzen mit dem Programm „Pregap4“ für die Assemblierung vorbereitet. Anschließend erfolgte das Zusammenfügen der Rohsequenzen mit dem Programmpaket „Gap4“. Die manuelle Nachbearbeitung der so erhaltenen Ergebnisse führte beim Klon 98L15 zu neun und beim Klon 49E3 zu sieben größeren zusammenhängenden Sequenzbruchstücken (Contigs).

Im zweiten Schritt folgte das Schließen der Lücken zwischen den Contigs. Dazu wurden jeweils an den Enden der Contigs Primer geplant, um dann durch Sequenzierung die Enden der Contigs zu verlängern und letztendlich die Lücken zu schließen („primer walking“). Das „primer walking“ erfolgte direkt auf der PAC-DNA nach einem speziellen Protokoll (Kapitel 4.8.3) für die Sequenzierreaktion. Ferner wurde die im Rahmen des

Human-Genom-Projekts geforderte Abgabequalität der Sequenz sichergestellt. Die Kriterien hierfür lauten:

1. Alle Bereiche der zusammengefügt Sequenz müssen mindestens durch drei Rohsequenzen abgedeckt sein.
2. Für jede Base der zusammengefügt Sequenz wird durch ein Programm (Phred) ein Qualitätswert zugeordnet. Dieser wird aus den Rohsequenzen ermittelt. Die Abgabequalität ist dann erreicht wenn der Wert 20 nicht unterschritten wird.

Um die Abgabequalität der Sequenz des Klons 98L15 zu erzielen wurde folgendermaßen vorgegangen. Wenn möglich, wurde die Sequenz einzelner „shotgun“-Klone durch „primer walking“ verlängert. Andererseits erfolgte das „primer walking“ auf der PAC-DNA dieses Klons. Die Durchführung dieser Sequenzierungen erfolgte nach ABI-Protokoll (Kapitel 4.8.3). Bei dem Klon 98L15 waren für das Schließen der Lücken und für das Erreichen der Abgabequalität insgesamt 29 Primer notwendig. Die hierzu benötigten Primer sind im Anhang B (Kapitel 7.2.1) aufgelistet. Die Länge der Sequenz des fertig gestellten Klons beträgt 116189 bp und wurde aus 1502 Sequenzierungen mit einer daraus folgenden 8,28-fachen Abdeckung erhalten.

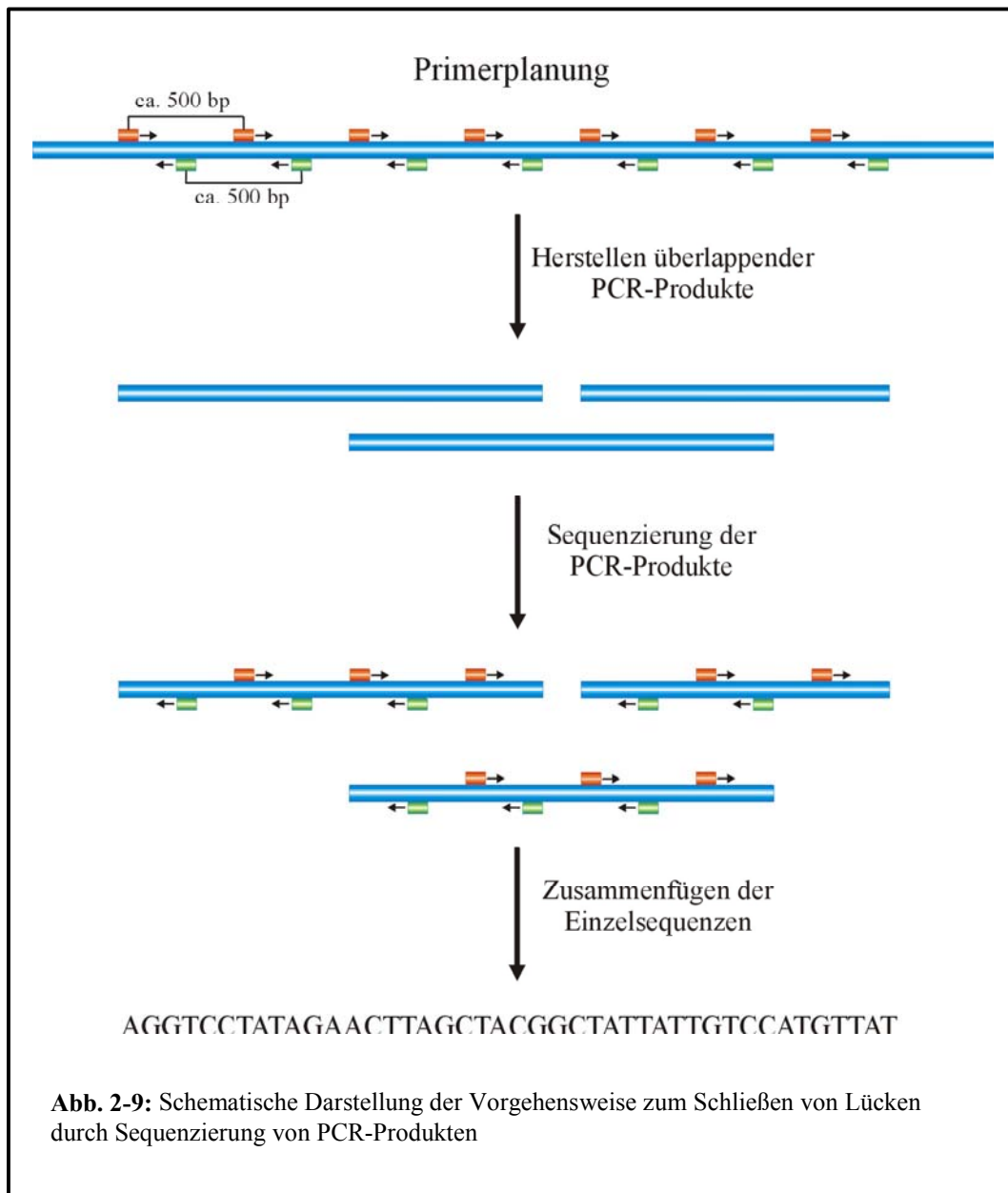
Im Fall des Klons 49E3 gestaltete sich das Schließen der Sequenzlücken mittels „primer walking“ wesentlich schwieriger. Das „primer walking“ auf der DNA dieses Klons führte zu schlechten Sequenzierungsergebnissen, die nicht verwendet werden konnten. Hilfreich war in dieser Situation, dass der Klon R73i15 bereits vom Partner-Institut RIKEN analysiert worden war. Laut Kartierungsdaten überlappt dieser Klon auf einer Länge von ca. 90 kbp mit dem Klon 49E3. Dadurch war es möglich, die Reihenfolge der 7 Contigs des Klons 49E3 zu ermitteln und die Größe der Lücken zwischen den Contigs zu bestimmen. Die Contigs wurden für die weitere Bearbeitung der Einfachheit halber entsprechend ihrer Abfolge mit den Buchstaben A bis G bezeichnet.

Lücke zwischen Contig:

A und B :	1051 bp
B und C:	2581 bp
C und D:	1363 bp
D und E:	120 bp
E und F:	8467 bp
F und G:	2124 bp

Innerhalb der Sequenzlücke zwischen den Contigs E und F konnte durch Sequenzvergleiche mit dem Klon R73i15 noch ein 914 bp großes Contig angeordnet werden. Dieses Contig endet 37 bp vor Beginn des Contigs F. Daraus ergibt sich eine fehlende Sequenz von 14792 bp Länge. Aufgrund der bekannten Sequenz innerhalb dieser Lücken bot sich die Möglichkeit durch entsprechend gewählte Primer Sequenzbereiche mittels PCR zu amplifizieren und anschließend die aufgereinigten Produkte in die Sequenzierreaktion einzusetzen (Abb. 2-9). Dazu wurden anhand der Sequenz des Klons R73i15 auf dem Plusstrang und auf dem Minusstrang jeweils Primer in einem Abstand von 400-500 bp geplant. Ein Primer auf dem Minusstrang wurde mittig zwischen zwei Primern auf dem Plusstrang gesetzt. Anschließend wurden Primerpaare ausgewählt, mit denen 1-2 kbp große PCR-Produkte erhalten lassen, die sich jeweils an den Enden überlappen. Die PCR wurde wie in Kapitel 5.7.1 beschrieben durchgeführt. Anstelle von transformierten Bakterienzellen wurden allerdings 25 ng DNA des Klons 49E3 in die Reaktion eingesetzt. Generell wurde in die PCR-Reaktion eine Standard-*Taq*-Polymerase ohne „Proof-reading“-Aktivität eingesetzt. Nur in Einzelfällen wurde die Vent-Polymerase mit „Proof-reading“-Aktivität verwendet. Die Konzentrationen der aufgereinigten PCR-Produkte wurde UV-spektrometrisch (Kapitel 4.4.8) bestimmt und anschließend 90-120 ng des PCR-Produkts in die Sequenzierreaktion nach ABI-Protokoll (Kapitel 4.8.3) eingesetzt. Als Sequenzierprimer wurden jeweils die zuvor geplanten Primer verwendet, die innerhalb des jeweiligen PCR-Produkts binden. Anschließend folgte die Sequenzierung auf dem ABI377 entsprechend der in Kapitel 4.9 beschriebenen Protokolle. Mit dieser Vorgehensweise war es möglich, sechs Sequenzlücken vollständig zu schließen. Einzig zwischen den Contigs E und F verblieb eine Sequenzlücke mit einer Größe von 564 bp. Von diesem Bereich konnte kein verwertbares PCR-Produkt erhalten werden, da überwiegend Nebenprodukte gebildet wurden. Auch der Versuch, diese Lücke durch „primer walking“ auf der PAC-DNA nach ABI-Protokoll (Kapitel 4.8.3) zu schließen, schlug

fehl. Erst die Wiederholung der Sequenzierung mit einem modifizierten LICOR-Protokoll (Kapitel 4.8.3) führte zum Erfolg.



Um die Abgabequalität der Sequenz zu erzielen, wurde auf unterschiedliche Weise vorgegangen. Wenn möglich wurde die Sequenz einzelner „shotgun“-Klone durch „primer walking“ verlängert. Andererseits wurde die DNA des Klons 49E3 für die Sequenzierung verwendet. Führt diese Vorgehensweise nicht zum Erfolg, wurden wieder PCR-Produkte (Kapitel 4.8.1) erzeugt und anschließend in die Sequenzierungsreaktion eingesetzt. Die Durchführung dieser Sequenzierungen erfolgte nach ABI-Protokoll (Kapitel 4.8.3). Für die Qualitätssicherung des 8467 bp großen Bereichs zwischen den Contigs E und F wurde eine alternative Methode ein-

gesetzt. Dieser Bereich wurde vorwiegend durch die Sequenzierung von PCR-Produkten geschlossen und lag zu großen Teilen nur zweifach abgedeckt vor. Um die aufwendige Herstellung von PCR-Produkten zu vermeiden, wurde deshalb nach Restriktionsenzymen gesucht, mit denen Fragmente erhalten werden können, die diesen Bereich lückenlos abdecken und nur wenig über diesen hinausgehen. Ferner sollten die Fragmente eine gewisse Mindestgröße von ca. 1 kbp nicht unterschreiten, im Agarosegel leicht identifizierbar und dadurch leicht isolierbar sein. Für diesen Zweck stellte sich die Kombination der Restriktionsenzyme *Bgl* II und *Not* I als beste Kombination heraus. Mit diesen Restriktionsenzymen war es möglich den Bereich von Position 73661-86185 mit drei Fragmenten lückenlos abzudecken. Ausgehend von der Position 73661 wies das erste Fragment eine Größe von 5304 bp, das zweite eine Größe von 1318 bp und das dritte eine Größe von 5903 bp auf. Für den Restriktionsverdau (Kapitel 4.6.1) wurden 10µg DNA des Klon 49E3 eingesetzt. Anschließend wurden die Proben auf ein Agarosegel aufgetragen und elektrophoretisch aufgetrennt (Kapitel 4.9.1). Die Fragmente wurden aus dem Agarosegel isoliert (Kapitel 4.5.5) und einzelsträngige Bereiche an den Enden der Fragmente enzymatisch aufgefüllt (Kapitel 4.6.3.2). Danach wurden die Fragmente in den Vektor pUC18R ligiert (Kapitel 4.6.5) und transformiert (Kapitel 4.4.4). Von jeweils 5 Klonen eines Fragments wurde im analytischen Maßstab DNA präpariert (Kapitel 4.5.1). Diese wurden mittels Restriktionsverdau (Kapitel 4.6.1) mit dem Enzym *EcoR* I und anschließender gelelektrophoretischer Auftrennung (Kapitel 4.9.1) überprüft. Die präparierte DNA eines Klon jedes Fragments wurde für die Sequenzierung eingesetzt. Diese erfolgte auf dem ABI377 unter Verwendung des Standard-Sequenzierprotokolls (Kapitel 4.8.3). Als Sequenzierprimer wurden wieder die für die Sequenzierung auf PCR-Produkten verwendeten Primer eingesetzt.

Bei dem Klon 49E3 waren für das Schließen der Lücken und für das Erreichen der Abgabqualität insgesamt 174 Primer notwendig. Die hierzu benötigten Primer sind im Anhang B (Kapitel 7.2.2) aufgelistet. Die Länge der Sequenz des so fertig gestellten Klon betrug 105570 bp und wurde aus 2742 Sequenzierungen, mit einer daraus folgenden 11,29-fachen Abdeckung erhalten.

Nach Abschluss der Sequenzierarbeiten wurden die Bereiche der Sequenzüberlappung der beiden Klone 98L15 und 49E3 und von den jeweils daran anschließenden Klonen ermittelt. Laut Kartierungsdaten liegt folgende Reihenfolge der Klonen vor:

R73i15 (RIKEN) - 49E3 - 98L15 - 90B5 (IMB-Jena)

Zur Ermittlung der Länge der Sequenzüberlappung zwischen diesen Klonen wurde das Programm „overlap“ (Hornischer, GBF) verwendet. Die Ergebnisse aus diesem Sequenzvergleich sind in der Tabelle 2-14 aufgeführt.

Tab. 2-14: Ergebnisse aus den Sequenzvergleichen zur Ermittlung der Länge von Überlappungsbereichen benachbarter Klone.

Überlappung zwischen den Klonen	Länge der Überlappung (bp)
R73i15 (RIKEN) – 49E3	88910
49E3 – 98L15	17757
98L15 – 90B5 (IMB-Jena)	89801

In den ermittelten Überlappungsbereichen der jeweiligen Klone wurde nach Sequenzunterschieden gesucht. Die Klone 49E3, 98L15 und 90B5 (IMB-Jena) stammen aus der gleichen Bibliothek (RPCI-1, 3-5 library). Im Überlappungsbereich der Klone 49E3 und 98L15 wurden keine Sequenzunterschiede gefunden. Hingegen wurde innerhalb des Überlappungsbereichs der Klone 98L15 und 90B5 (IMB-Jena) eine Deletion gefunden. In der Sequenz des Klons 90B5 wurden 12 aufeinander folgende T's gefunden, an deren Stelle auf dem Klon 98L15 hingegen 13 aufeinander folgende T's zu finden sind. Diese Deletion ließ sich nicht eindeutig auf einen Editierungsfehler zurückführen. Da in diesem Bereich die Consensussequenz des Klons 98L15 durch eine größere Anzahl „shotgun“-Sequenzen abgedeckt ist als der entsprechende Bereich auf dem Klon 90B5, wurden die 13 T's in die Consensussequenz des Chromosoms 21 übernommen. Die genaue Angabe zur Position der Deletion findet sich im Anhang C (Kapitel 7.3.1). Die Klone 49E3 und R73i15 stammen aus unterschiedlichen Bibliotheken. Der Klon R73i15 stammt aus der Bibliothek RPCI-11. Beim Vergleich der Sequenzen dieser beiden Klone konnten innerhalb des Überlappungsbereichs 31 Deletionen, 208 Einzelbasenaustausche und 3 Mehrfachbasenaustausche gefunden werden. Die Sequenz des Klons 49E3 wurde schon auf mögliche Editierungsfehler untersucht. Von dieser Seite sind die gefundenen Sequenzunterschiede bestätigt. Die genauen Angaben zu den Deletionen und Basenaustauschen finden sich im Anhang C (Kapitel 7.3.2). Die Ergebnisse des Sequenzvergleichs wurden zu „RIKEN“ in Japan gesendet, um die Sequenz des Klons R73i15 auf Editierungsfehler zu untersuchen. Das Ergebnis dieser Untersuchung lag zum Zeitpunkt der Fertigstellung dieser Arbeit noch nicht vor.

Anschließend wurden die Sequenzen der Klone 49E3 und 98L15 annotiert, um so mögliche Gene vorauszusagen oder bekannte Gene aufzufinden. Die Durchführung der „first-pass“-Annotierung und die anschließende Auswertung der erhaltenen Daten wird in Kapitel 4.12 beschrieben. Die Auswertung der „first-pass“-Annotierung des Klons 49E3 und des Klons 98L15 erfolgte nach Kriterien, die vom Chromosom 21 Konsortium festgelegt wurden (Kapitel 4.12). Im Anhang D (Kapitel 7.4) ist am Beispiel des Klons 98L15 ein Auszug der „first-pass“-Annotierung vom Dezember 2001 angegeben, die mit dem Programm „AnnoMitter“ erzeugt wurde. Ferner findet sich dort (Kapitel 7.4.1) die graphische Darstellung dieser Annotierungsdaten, die mit dem Programm „showseq“ (Hornischer, GBF) erzeugt wurde, um so die Daten einfacher bearbeiten zu können.

Unter Anwendung der festgelegten Kriterien konnte auf dem Klon 49E3 mit dem momentanen Stand der EST-Datenbanken (Dezember 2001) und den vorliegenden Versionen der Exonvorhersageprogramme kein Gen lokalisiert werden.

Durch die Auswertung der Annotierungsergebnisse des Klons 98L15 konnten so zwei mögliche Gene gefunden werden. Das erste Gen liegt in Minusrichtung an der Position 12742-34918 und weist eine Länge von 25177 bp auf. Dieses Gen ist unvollständig, sowohl Anfang als auch Ende des Gens konnten nicht gefunden werden. In diesem Fall konnte eine EST-Sequenz gefunden werden, die sich über alle 4 Exons des Gens erstreckt und zu annähernd 100% mit der Sequenz der genomischen DNA übereinstimmt. Das Gen setzt sich wie folgt aus vier Exons zusammen:

Exon Nr. 4	Position	12742-12843
Exon Nr. 3	Position	16945-17053
Exon Nr. 2	Position	20018-20403
Exon Nr. 1	Position	37800-37918

Durch zusammenfügen der Sequenzen dieser Exons ergab sich folgende Nukleotidsequenz:

```

1   GGTCTTATTT ATGTTGCTCA AATGTTTACA TAATAACTTG TGCCAAAAAT ATAGTGCTCA  60
61  TGCTTTTTCAG TTCTCACCCA GAAATGTCCT GTGGCTTCTA GTTGTGTGCC TGAGATCAGA 120
121 TAATAAAAGA CCATGCCTTG AATTCTCTCA GCTAAGTGTA AAGGATTCCT TCAGAGATTT 180
181 ATTTATTCCG AGAATAGAGA CCATTCTGAT GATGTATACA AGGAACAACC TAAACTGTGC 240
241 TGAGCCACTG TTTGAACAAA ATAACACTACT TAATGTTAAT TTCAACACAC AAAAGAAAAAC 300
301 AGTCTGGCTT ATTCACGGAT ACAGACCAGT AGGCTCCATC CCATTATGGC TTCAGAACTT 360
361 CGTAAGGATT TTGCTGAATG AAGAAGATAT GAATGTAATT GTAGTAGACT GGAGCCGGGG 420
421 TGCTACAAC TTTATTTATA ATAGAGCAGT TAAAAACACC AGAAAAGTTG CTGTGAGTTT 480
481 GAGTGTGCAC ATTAAAAATC TTTTGAAGCA TGGTGCATCT CTTGACAATT TTCATTTTCAT 540
541 AGGTGTGAGC TTAGGGGCTC ATATCAGTGG ATTTGTTGGA AAGATATTTT ATGGTCAACT 600
601 TGGAAGAATA ACAGGTCTTG ACCCTGCTGG GCCAAGGTTT TCCAGAAAAAC CACCATATAG 660
661 CAGATTAGAT TACACGGATG CAAAGTTTGT GGATGTCATC CATTCTGACT CCAATG      716

```

Die Übersetzung dieser Nukleotidsequenz in die Aminosäuren ergab im Leseraster 3 die folgende Sequenz:

```

1   VLFMLLKCLH NNLCQKYS AH AFQFSPRNV L WLLVVCLRSD NKRPCLEFSQ LSVKDSFRDL  60
61  FIPRIETILM MYTRNNLNCA EPLFEQNNSL NVNFNTQKKT VWLIHGYPV GSIPLWLQNF 120
121 VRILLNEEDM NVIVVDWSRG ATTFIYNRAV KNTRKVAVSL SVHIKNLLKH GASLDNFHFI 180
181 GVS LGAHISG FVGKIFHGQL GRITGLDPAG PRFSRKPPYS RLDYTDK FV DVIHSDSNX  239

```

Der Vergleich dieser Sequenz mit der Proteindatenbank „pfam“ (Kapitel 5.11) ergab eine signifikante Übereinstimmung mit der Triglycerid Lipase (EC 3.1.1.3).

Das zweite Gen liegt in Plusrichtung an der Position 47051-58371 und weist eine Länge von 11321 bp auf. In diesem Fall konnte eine EST-Sequenz gefunden werden, die sich über alle 5 Exons des Gens erstreckt und zu annähernd 100% mit der Sequenz der genomischen DNA übereinstimmt. Die so gefundenen Exons 1,2 und 5 konnten zusätzlich durch die Ergebnisse der Vorhersageprogramme bestätigt werden. Dieses Gen ist vollständig und setzt sich folgendermaßen zusammen:

Ergebnisse

5'-UTR	Position 47051-47171
Exon Nr.1	Position 47051-47267
Exon Nr. 2	Position 50548-50710
Exon Nr. 3	Position 52082-52154
Exon Nr.4	Position 55423-55522
Exon Nr.5	Position 57865-58371
3'-UTR	Position 58279-58371
Poly A-Signal	Position 58352-58357

Durch zusammenfügen der Exons ergab sich folgende Nukleotidsequenz:

```

1      ATGCCTCGCC CCCAGCGCGC GCTCCCGGCG CCGGCCCCGC CCCCTCACCC AGGCGGCGCC      60
61     GTAAAGCCGG AGAAGGGGCG GGGTCTCAGC TCCTACTTCA TTCTACGGCC GAGACCGGAG      120
121    GATGTTCCCT GCTCAGGAGG AGGCCGACAG GACCGTGTTT GTTGGGAATT TAGAGGCCCG      180
181    AGTTCGGGAA GAGATTCTGT ACGAGCTGTT CCTTCAGGCG GGGCCACTAA CCAAAGTGAC      240
241    TATATGCAAA GACAGAGAAG GAAAGCCAAA GTCTTTTGGA TTTGTCTGCT TTAAACACCC      300
301    AGAATCGGTG TCTTATGCCA TAGCTTTGCT GAATGGAATT CGTTTATATG GAAGACCAAT      360
361    TAACGTGCAG TATCGATTTG GGAGTTCTCG CTCTTCTGAA CCAGCTAACC AAAGTTTGA      420
421    GAGCTGTGTT AAGATAAATT CACACAATA CAGGAATGAA GAAATGTTGG TGGGCAGATC      480
481    TTCCTTTCCC ATGCAGTATT TTCCAATTAA TAATACTTCT TTACCTCAAG AATATTTTCT      540
541    CTTTCAGAAG ATGCAGTGGC ATGTGTATAA TCCAGTGCTG CAGCTTCCTT ACTATGAAAT      600
601    GACAGCTCCA CTTCCTAATA GTGCATCCGT GTCTTCCTCA CTGAATCATG TTCCAGATCT      660
661    TGAGGCTGGA CCCAGCTCAT ATAAATGGAC TCACCAACAA CCAAGTGACT CTGACCTTTA      720
721    TCAGATGACA GCTCCACTTC CTAATAGTGC ATCCGTGTCT TCCTCACTGA ATCATGTTCC      780
781    AGATCTTGAG GCTGGACCCA GCTCATATAA ATGGACTCAC CAACAACCAA GTGACTCTGA      840
841    CCTTTATCAG ATGAATAAAC GAAAGAGACA AAAGCAAACA AGTGATAGTG ATAGTAGCAC      900
901    AGACAACAAC AGAGGCAACG AATGTAGCCA AAAGTTCCGA AAGTCTAAGA AGAAGAAAAG      960
961    ATACTAGTAT TACCTACAAA TGAACTTAC CTACACTGAT CTTAGTTCTC TTATGAAAAA      1020
1021   AATAAGATGT TATCCCATCA AATAACAAT GTCATGGCT

```

Die Übersetzung dieser Nukleotidsequenz in die Aminosäuren ergab im Leseraster 3 die folgende Sequenz:

```

1  CLAPSARSRR RPRPLTQAAP *SRRRGGVSA PTSFYGRDRR MFPAQEEADR TVFVGNLEAR 60
61  VREEILYELF LQAGPLTKVT ICKDREGKPK SFGFVCFKHP ESVSYAIALL NGIRLYGRPI 120
121 NVQYRFGSSR SSEPANQSFE SCVKINSHNY RNEEMLVGRS SFPMQYFPIN NTSLPQEYFL 180
181 FQKMQWHVYN PVLQLPYYEM TAPLPNSASV SSSLNHVPDL EAGPSSYKWT HQQPSDSDLY 240
241 QMTAPLPNSA SVSSSLNHVP DLEAGPSSYK WTHQQPSDSD LYQMNRKRQ KQTSDDSDSST 300
301 DNNRGNECSQ KFRKSKKKKR Y*YYLQMKLT YTDLSSLMKK IRCYPIK*TM SWX 353

```

Der rot markierte Bereich entspricht der 5'-UTR bzw. der 3'-UTR. Entsprechend wurde nur die dazwischen liegende Aminosäuresequenz für den Vergleich mit der Proteindatenbank „pfam“ (Kapitel 4.12) herangezogen. Dieser Vergleich ergab eine signifikante Übereinstimmung mit einem „RNA binding motif protein 11“.

Schon zum Zeitpunkt der Fertigstellung des Klons 98L15 im Juni 1999 wurde die Sequenz erstmals annotiert. Die beiden möglichen Gene wurden in den Genkatalog von Chromosom 21 aufgenommen (The chromosome 21 mapping and sequencing consortium, 2000). Das Gen 2 („RNA binding motif protein 11“) konnte schon zu diesem Zeitpunkt vollständig gefunden werden. Im Fall von Gen 1 (Triglycerid Lipase) waren zum damaligen Zeitpunkt 3 Exons dieses Gens gefunden worden. Unter Verwendung der aktuelleren EST-Datenbank war es möglich dieses Gen um ein weiteres Exon (Exon 1) zu verlängern. Die vollständigen Sequenzen und „first-pass“-Annotierungsdaten der Klone 98L15 und 49E3 sind über das Internet unter der Adresse http://genome.gbf.de/index_d.html abrufbar.

2.5 Entwicklung einer Vorrichtung zur manuellen Übertragung von Zellkolonien in Mikrotiterplatten

Um Zellkolonien von Agarplatten in Mikrotiterplatten zu übertragen werden entweder sterile Zahnstocher oder Roboter verwendet. Die Verwendung von sterilen Zahnstochern ist sehr kostengünstig, aber auch sehr langsam. Das Animpfen einer Mikrotiterplatte im 384er Format dauert mit dieser Methode ca. 1 Stunde. Ferner kommt es häufig zu Fehlern, wie nicht erfasste Positionen oder doppelte Animpfungen in der Mikrotiterplatte. Bei der sehr exakt und schnell arbeitenden, rechnergestützten Robotertechnologie werden bildverarbeitende Systeme verwendet. Ein per CCD-Kamera aufgenommenes Bild der Agarplatte mit den darauf angewachsenen Kolonien wird mittels digitaler Bildanalyse ausgewertet. Durch Auswahlparameter wie

Größe, Form, Farbe und Abstand zu benachbarten Kolonien kann sehr genau bestimmt werden welche Kolonien gepickt werden sollen. Die Geschwindigkeit mit der Zellkolonien übertragen werden können variiert je nach System von 2000 bis 3500 Kolonien pro Stunde. Robotersysteme kosten allerdings, je nach Leistungsfähigkeit 50000-250000 Euro und amortisieren sich daher erst bei einem hohen Durchsatz über lange Zeit.

Sind Klonierungsexperimente zu umfangreich, um mit der Zahnstocher-Methode bewältigt werden zu können, ist der Einsatz eines rechnergestützten Robotersystems unumgänglich. Derartige Robotersysteme sind insbesondere für Hochdurchsatzprojekte, wie z. B. das Humangenomprojekt sinnvoll. Für weniger umfangreiche Projekte sind diese Systeme zu teuer. Für diese Fälle wurde im Rahmen dieser Arbeit eine preiswerte und in der Handhabung einfache Vorrichtung (im folgenden als „Picking-Tool“ bezeichnet) für die Übertragung von Zellkolonien geschaffen.

Folgende Überlegungen führten zur Konstruktion des Picking-Tools. Die Zahnstocher-Methode ist aus den folgenden Gründen so zeitaufwendig. Es kann immer nur eine Zellkolonie aufgenommen und anschließend in die Mikrotiterplatte übertragen werden. Bei der Übertragung einer Zellkolonie ist darauf zu achten die richtige Position in der Mikrotiterplatte zu beimpfen. Dies ist gerade bei einer Mikrotiterplatte im 384er Format sehr schwierig. Entsprechend ist es zur Beschleunigung des manuellen Pickens von Zellkolonien erforderlich, mehrere Kolonien nacheinander einzeln aufzunehmen und diese dann in einem Arbeitsgang in die Gefäße einer Mikrotiterplatte zu übertragen. Optimal war es, so viele Kolonien auf einmal aufzunehmen, wie eine Reihe der jeweils vorliegenden Mikrotiterplatte fassen kann. Bei einer Mikrotiterplatte im 384er Format müssen entsprechend pro Reihe 24 Kolonien aufgenommen und übertragen werden. Die Abbildungen 2-10 und 2-11 zeigen das Picking-Tools der ersten Generation. Das Picking-Tool besteht aus zwei Komponenten. Einerseits aus einem speziell geformten Kamm, in dessen Längsrichtung eine Transportperforation verläuft und andererseits aus einer Vorrichtung zur Aufnahme dieser Kämme. Diese Vorrichtung verfügt über eine Mechanik mit der der Kamm jeweils um eine Position weiter transportiert wird. Durch eine Gleitkufe wird jeweils ein Zahn nach unten gebogen und kann so mit Zellen benetzt werden, ohne dass die benachbarten Zähne kontaminiert werden. Für die Herstellung dieser Kämme wurde V2A-Folie verwendet. Die Elastizität des Materials erlaubt ein relativ starkes Verbiegen mit anschließender Rückfederung in die Ursprungslage. Ferner besitzt dieses Material eine hohe Haltbarkeit und bietet die Möglichkeit einer einfachen Sterilisation durch Autoklavieren.



Abb. 2-10: Gezeigt ist das Pickingtool der ersten Entwicklungsstufe im Einsatz.

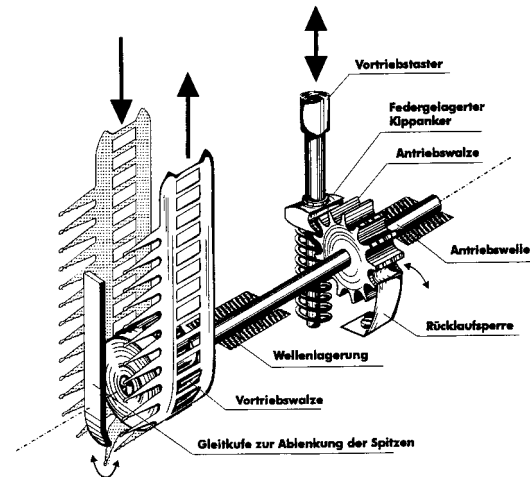


Abb. 2-11: Darstellung der Mechanik des Pickingtools der Entwicklungsstufe 1.

Das Picking-Tool wurde im praktischen Einsatz auf seine Funktionalität überprüft. Dazu wurden Bakterienkolonien von Agarplatten in Mikrotiterplatten im 384er Format übertragen. Mit dem Picking-Tool konnten pro Stunde vier Mikrotiterplatten dieses Formats fehlerfrei beimpft werden. Im Vergleich zur Zahnstocher-Methode lässt sich mit dem Picking-Tool das manuelle Übertragen von Zellkolonien in Mikrotiterplatten viermal schneller verrichten. Die mit dem Picking-Tool erzielbare Übertragungsgeschwindigkeit von 1536 Zellkolonien pro Stunde reicht sogar an Robotersysteme heran.

Für die Weiterentwicklung des Picking-Tools wurden die Erkenntnisse aus dem Laboreinsatz der ersten Version zugrundegelegt. Die Überarbeitung galt vorwiegend der Mechanik und der Kämme. Ziel dieser Maßnahme war es, die Störanfälligkeit der Mechanik zu minimieren. Zu diesem Zweck wurde die Mechanik durch Reduzierung der beweglichen Teile vereinfacht. Dadurch konnte diese zusätzlich wesentlich kompakter gestaltet werden. Die Funktionsweise der geänderten Mechanik erforderte eine Modifikation der Kämme. Entlang der Transportperforation der Kämme wurde durch Prägung eine Wölbung erzeugt. Um die Formstabilität der Kämme zu erhöhen, wurden die Aussenkanten der Zähne geprägt. Dadurch war das Kammgehäuse der ersten Version des Picking-Tools zur Führung des Kamms nicht mehr notwendig und konnte somit entfallen. Die Abbildungen 2-12 und 2-13 zeigen das Picking-Tool der zweiten Generation.



Abb. 2-12: Pickingtool der Entwicklungsstufe 2 im Laboreinsatz

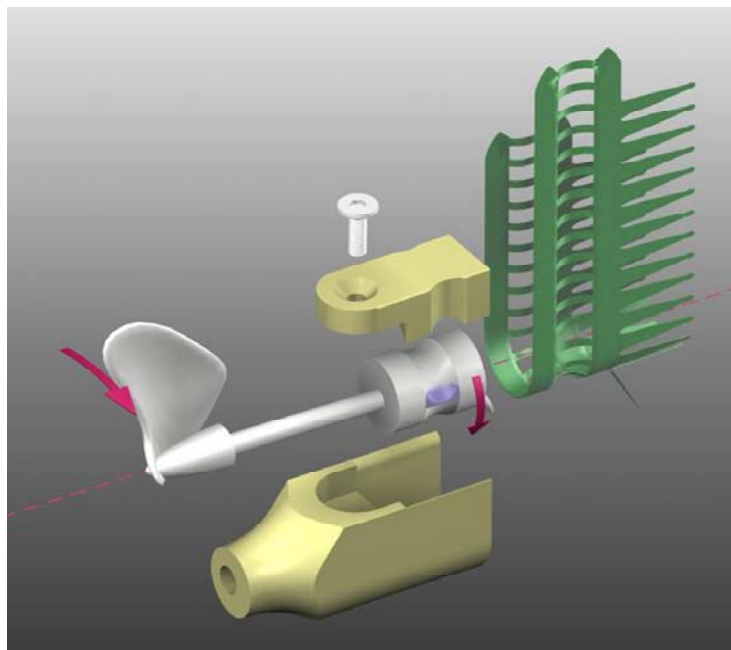


Abb. 2-13: Darstellung der Mechanik des Pickingtools der zweiten Entwicklungsstufe.

In Langzeittests arbeitete die verbesserte Mechanik fehlerfrei. Ferner führte die durch Prägung verbesserte Formstabilität zu einer deutlich erhöhten Haltbarkeit der Kämmen. Die ergonomisch noch nicht optimale Handhabung dieser zweiten Version des Picking-Tools führte zur Revision des Designs. Diese verbesserte dritte Version ist in der Zeichnung in Abb. 2-14 dargestellt.

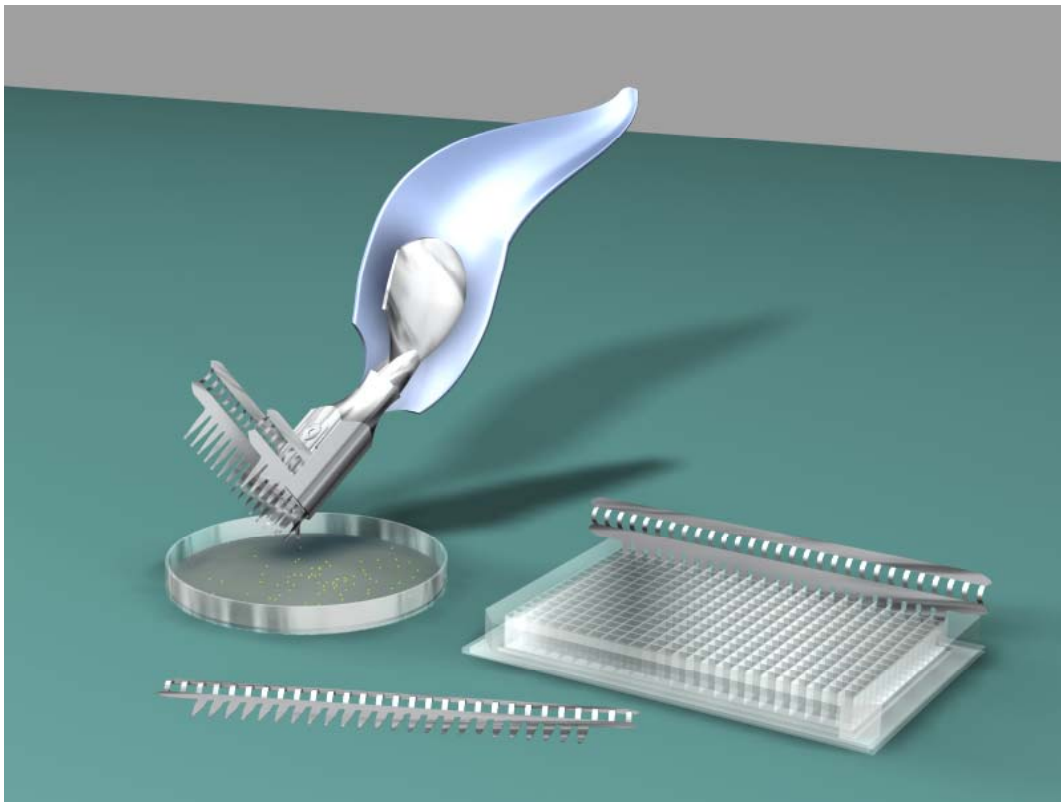


Abb. 2-14: Entwurfsskizze des Pickingtools der Entwicklungsstufe 3.

Die notwendigen Änderungen galten der Gestaltung des Griffteils. Der Griff entspricht nun einer Art Schale, die optimal an die Anatomie der Hand angepasst wurde. Die Haltung des Tools während der Arbeit ist einem Schreibstift ähnlich. Sollten sich die Ergonomieverbesserungen im praktischen Langzeiteinsatz bestätigen, wäre das Picking-Tool in seiner dritten Version ein Labormuster, das zur Industriereife gebracht wurde.

3 Diskussion

3.1 Optimierte "shotgun"-Banken-Herstellung

Das Herstellungsverfahren von „shotgun“-Banken (Kapitel 2.1) konnte durch Optimierung der Einzelschritte deutlich verbessert werden. Mit dem optimierten DNA-Präparationsprotokoll (Kapitel 2.1.1.2.4) konnte der Anteil an chromosomaler DNA aus *E. coli* auf einen durchschnittlichen Wert von 3,1% gesenkt werden. Im Vergleich zum anfänglich verwendeten Präparationsprotokoll (Kapitel 2.1.1.1) wurde dieser Anteil um einen Faktor von ca. 4,4 gesenkt. Durch Optimierung des gesamten Herstellungsverfahrens kann nun die sechsfache Anzahl an „shotgun“-Banken im gleichen Zeitraum (Kapitel 2.1.4) hergestellt werden. Zusätzlich konnte die notwendige Anzahl an Versuchen für die erfolgreiche Herstellung einer „shotgun“-Bank um etwa den Faktor 3 gesenkt werden. Mit dem optimierten Verfahren sind durchschnittlich 1,4 Versuche für die erfolgreiche Herstellung einer „shotgun“-Bank notwendig. Im Vergleich zum ursprünglichen Herstellungsverfahren lässt sich so etwa die 18-fache Anzahl von „shotgun“-Banken in gleichen Zeitraum herstellen. Ab Mitte des Jahres 1997 nahm die Abteilung Genomanalyse der GBF an dem Human-Genom-Projekt teil, mit der Option innerhalb von 3 Jahren sechs Megabasen Sequenz des Chromosoms 21 zu analysieren. Dazu mussten im ersten Jahr eine, im zweiten Jahr zwei und im dritten Jahr drei Megabasen Sequenz fertig gestellt werden. Für die erste Megabase Sequenz waren ca. acht "shotgun"-Banken notwendig. Aufgrund der häufigen Fehlversuche, der Dauer und der geringen Parallelisierbarkeit beim anfänglich verwendeten Herstellungsverfahren bildete die Fertigstellung dieser acht "shotgun"-Banken den limitierenden Faktor zur Erreichung des ersten Jahresziels. Das letztendlich eingesetzte, optimierte Herstellungsverfahren bietet die Möglichkeit, dass pro Jahr von einer Person eine ausreichende Menge an "shotgun"-Banken hergestellt werden kann, die zur Analyse von ca. 15-20 Megabasen Sequenz notwendig ist.

Das anfänglich eingesetzte Herstellungsverfahren wurde einer Promotionsarbeit (P. Scholler, 1996) entnommen, in der unter anderem die Etablierung eines neuen Verfahrens zur Entwicklung und Signaldetektion von Hybridisierungsfiltern für die Kartierung von "shotgun"-Banken untersucht wurde. Die Anforderungen an "shotgun"-Banken, die für Kartierungsarbeiten bestimmt sind, unterscheiden sich von denen, die an "shotgun"-Banken gestellt werden, die in die Sequenzierung eingesetzt werden. Einerseits stellt die Kartierung selbst den zeitlich limitierenden Faktor dar, d. h. es werden im gleichen Zeitraum von vornherein weniger "shotgun"-Banken benötigt. Andererseits ist der Anteil von Klonen, die als Insertion im Klonie-

rungsvektor Bruchstücke chromosomaler Bakterien-DNA tragen, in einer solchen "shotgun"-Bank von untergeordneter Bedeutung, da diese während der Kartierung ausselektioniert werden. Dies bedeutet für die Herstellung solcher Banken, dass die Abtrennung chromosomaler Bakterien-DNA während der DNA-Präparation nicht das vorrangige Ziel darstellt. Vielmehr war das Verfahren im Hinblick auf eine große Menge an DNA optimiert, um daraus eine "shotgun"-Bank für die Kartierung mit diesem Verfahren herstellen zu können.

Die anfänglichen Fehlversuche durch Fragmentierung der DNA mit Ultraschall sind (Kapitel 2.1.2.1) vor allem auf das zur Verfügung stehende Ultraschallgerät zurückzuführen. Für die Fragmentierung von DNA gibt es geeignetere Ultraschallgeräte, wie z. B. das Gerät XL2025 der Fa. Misonix mit zusätzlichem Becheraufsatz „Cup Horn“ (pers. Mitt. T. Schlüter, Lion Bioscience). Mit diesem Gerät können gut reproduzierbare Ergebnisse erzielt werden. Ferner erlaubt dieser Aufbau die gleichzeitige Fragmentierung mehrerer unterschiedlicher DNA-Proben. Dadurch ergibt sich ein Geschwindigkeitsvorteil gegenüber Verfahren, bei denen nur Probe für Probe abgearbeitet werden kann. Allerdings kann auch bei diesem System, trotz der besseren Reproduzierbarkeit, nicht auf eine Kontrolle des Fragmentierungsergebnisses verzichtet werden. Dazu wird ein Teil der Probe und ein Längenstandard auf ein Agarosegel aufgetragen und elektrophoretisch aufgetrennt. Erst danach können die Proben für die Herstellung von „shotgun“-Banken verwendet werden.

Beim Einsatz eines Hydroschersystems (Kapitel 2.1.2.2) relativiert sich allerdings der oben angesprochene Geschwindigkeitsvorteil. Zwar kann man mit diesem Hydroschersystem Proben nur nacheinander abarbeiten, jedoch ist nach erfolgter Eichung erfahrungsgemäß die Reproduzierbarkeit der Fragmentierungsergebnisse so gut, dass auf eine nachfolgende Kontrolle verzichtet werden kann. Die Probe kann sofort zur Herstellung von „shotgun“-Banken eingesetzt werden.

Gerade im Hinblick auf die Herstellung von "shotgun"-Banken für die Sequenzierung zeichnet die Fragmentierung von DNA durch hydrodynamisches Scheren einen wichtigen Vorteil gegenüber der Ultraschallfragmentierung. Die DNA-Fragmentierung mit Ultraschall führt zu einem überproportional häufigen Brechen der DNA in AT-reichen Regionen (Hengen P.N.; 1997). Dies führt bei langgestreckten AT-reichen Regionen unter Standardbedingungen verstärkt zu besonders kleinen Fragmenten, von denen bei der nachfolgenden Fragmentlängenselektion nur wenig oder keine mit erfasst werden. Für eine daraus erstellte "shotgun"-Bank bedeutet das, dass Klone aus diesem Bereich deutlich unterrepräsentiert oder gar nicht vorhanden sind. Sind keine Klone vorhanden, entsteht bei der Sequenzierung zwangsläufig eine Lücke in der Consensussequenz. Diese muss dann durch zeit- und kostenintensives di-

rektes Sequenzieren auf der Ausgangs-DNA mittels entsprechend geplanter Primer geschlossen werden, oder es müssen unter modifizierten Bedingungen neue Banken hergestellt werden.

Bei der Fragmentierung durch hydrodynamisches Scheren ist das sequenzabhängige Brechen der DNA deutlich geringer ausgeprägt. Dadurch erhält man Fragmente, die annähernd statistisch über den gesamten Bereich der Ausgangs-DNA verteilt sind (Oefner P.J. *et al.*, 1996; Thorstenson Y.R. *et al.*, 1998). Eine daraus hergestellte "shotgun"-Bank führt deshalb nach Sequenzierung nicht zu Lücken in der Consensussequenz, die auf eine unterrepräsentierte Anzahl von Klonen eines bestimmten Sequenzbereichs zurückzuführen sind. Deshalb müssen weniger Lücken durch direktes Sequenzieren auf der Ausgangs-DNA geschlossen werden. Auch der eventuelle zusätzliche Aufwand für das Herstellen von Banken unter nicht standardisierten Bedingungen entfällt.

Wie in den Ergebnissen (Kapitel 2.1.3) gezeigt, konnte durch Einsatz eines Hydroschersystems und durch die optimierte enzymatische Auffüllreaktion die, zur Herstellung von "shotgun"-Banken, notwendige DNA-Menge deutlich reduziert werden. Dadurch wurden weitere Modifikationen bei der DNA-Präparation möglich, die zu einer Verringerung der genomischen DNA der *E. coli*-Zellen an der Gesamtmenge der präparierten DNA führten. Dies ermöglichte es geringere Zellmengen in die Präparation einzusetzen. Bei gleichzeitiger Erhöhung der Puffervolumen für die alkalische SDS-Lyse (Kapitel 2.1.1) führt dies zu einer verringerten Kontamination der präparierten DNA mit chromosomaler Bakterien-DNA. Dies lässt sich durch das Prinzip der alkalischen Lyse erklären (Birnboim H.C. & Doly J., 1979).

Das Grundprinzip der alkalischen Extraktionsmethode basiert darauf, dass lineare DNA im Gegensatz zu superhelikaler DNA bei einem pH-Wert von 12-12,5 denaturiert vorliegt (Currier T.C. *et al.*; 1976). In der Bakterienzelle liegen sowohl die chromosomale DNA als auch die Vektor-DNA zirkulär vor. Allerdings kommt es in der stationären Wachstumsphase und beim Abernten der Zellen schon zu einem enzymatischen Abbau der chromosomalen DNA (Reinikainen *et al.*; 1989). Ferner ist davon auszugehen, dass bei der alkalischen Lyse während der Durchmischung nach Zugabe des NaOH/SDS-Puffers Scherkräfte auftreten, die ebenfalls zur Fragmentierung der chromosomalen DNA führen (Levy M.S. *et al.*, 1999). Als Folge liegt die chromosomale DNA überwiegend fragmentiert, also linear vor. Die Vektor-DNA bleibt aufgrund des erheblich geringeren Molekulargewichts weitestgehend unbeeinträchtigt und liegt deshalb überwiegend in der superhelikalen Konformation vor. Dieser Umstand wird zur Aufreinigung der Vektor-DNA ausgenutzt. Durch Zugabe des NaOH/SDS-

Puffers zu den vollständig resuspendierten Zellen erfolgt die Lyse der Zellmembran. Die linearisierte chromosomale DNA, deren Fragmente immer noch ein sehr hohes Molekulargewicht aufweisen, liegt unter diesen pH-Wert-Bedingungen selektiv denaturiert vor. Nach der Neutralisation durch Zugabe von Kaliumacetat-Puffer renaturiert der größte Teil der chromosomalen DNA in Form eines unlöslichen Netzwerks. Gleichzeitig bedingt die hohe Konzentration von Kaliumacetat die Präzipitation von Protein-SDS-Komplexen und hochmolekularer RNA. Deshalb können diese Verunreinigungen anschließend durch Zentrifugation abgetrennt werden. Die Vektor-DNA kann in der Folge durch Fällung oder mittels Anionentauschersäule aus dem Überstand isoliert werden.

Es hat sich allerdings gezeigt, dass es sich teilweise sehr schwierig gestaltet, die chromosomale DNA zu entfernen. Durch Einwirkung zu hoher Scherkräfte wird die chromosomale DNA entsprechend stark fragmentiert und es resultieren teilweise sehr kleine Fragmente. Diese werden nicht in das unlösliche Netzwerk integriert und verbleiben in Lösung (Durland R.H. and Eastman E.M., 1998; Prazeres D. *et al.*, 1998). Ferner kann man sich vorstellen, dass dieses Netzwerk selbst sehr scherempfindlich ist und durch Einwirken zu hoher Kräfte Teile aus diesem Netzwerk herausgebrochen werden und wieder in Lösung gehen.

Für die Herstellung von „shotgun“-Banken mit dem anfänglich verwendeten Verfahren wurden große DNA-Mengen benötigt. Dazu wurden Bakterienzellen aus 3 Liter Kulturvolumen abgeerntet und in die alkalische Lyse eingesetzt. Während der alkalischen Lyse wurden für diese große Zellmenge mit je 42 mL je Puffer nur verhältnismäßig kleine Volumina eingesetzt. Dies führte nach Zugabe des NaOH/SDS-Puffers zu den resuspendierten Zellen zu einer relativ viskosen Lösung, die sich nur schwer vollständig durchmischen ließ. Folglich musste dieses Durchmischen sehr intensiv erfolgen. Gleiches gilt für die nachfolgende Neutralisation mit Kaliumacetat-Puffer. Bei beiden Schritten führte die intensive Durchmischung offensichtlich zu entsprechend hohen Scherkräften, die eine zu starke Fragmentierung und damit teilweise relativ kleine Fragmente der chromosomalen DNA zur Folge hatten. Wie oben erwähnt, verbleiben diese „kleinen“ Fragmente in Lösung und führen zu einer Verunreinigung der Vektor-DNA. In der Folge wurde die eingesetzte Zellmenge im Verhältnis zu den verwendeten Puffervolumina schrittweise verringert. Dadurch wurde die Viskosität der Lösung, die nach Zugabe des NaOH/SDS-Puffers zu den resuspendierten Zellen erhalten wurde, gesenkt. Deshalb konnte die vollständige Durchmischung der Lösung sehr vorsichtig mit geringen Scherkräften erfolgen und führte entsprechend zu einem geringeren Anteil an „kleinen“ Fragmenten chromosomaler DNA. Aus diesem Grund wurde eine wesentlich bessere Abtrennung der chromosomalen DNA erzielt.

Allerdings kam es auch bei diesem Präparationsverfahren im Einzelfall zu einer unzureichenden Abtrennung der chromosomalen DNA. Wie erwähnt musste die Präparation des Klons 49E3 von Chromosom 21 acht mal wiederholt werden, bis die chromosomale DNA in ausreichendem Maß abgetrennt werden konnte (Kapitel 1.1.2.4). Die Überprüfung des Anteils chromosomaler DNA in einer "shotgun"-Bank erfolgte durch Sequenzierung von 96 Klonen dieser Bank und anschließendem Vergleich der erhaltenen Sequenzen mit einer *E. coli*-Datenbank. Die Zahl kommt dadurch zustande, weil maximal 96 Proben auf einem ABI377 in einem Lauf parallel sequenziert werden können. Die Problematik bei dieser Überprüfungsmethode liegt darin, dass 96 Klone keine statistische Verteilung repräsentieren. Folglich ist durch diese Methode höchstens eine Annäherung an den tatsächlichen Anteil an Klonen, die als Insertion im Vektor Fragmente der chromosomalen DNA der *E. coli*-Zellen tragen, gegeben. Gerade bei der Analyse der einzelnen Sequenzierläufe der vollständig sequenzierten "shotgun"-Bank des Klons 49E3 wurde dieser Misstand deutlich. Dabei wurden einerseits Sequenzierläufe gefunden, bei denen 15 bis 22 der 96 Klone Teile der chromosomalen DNA der *E. coli*-Zellen trugen, und bei denen der Grenzwert von 15% somit überschritten wurde. Andererseits kamen Sequenzierläufe vor, bei denen maximal ein Klon gefunden wurde, der Teile der chromosomalen DNA der *E. coli*-Zellen trug. Hingegen lagen nur 4 von 27 Sequenzierläufen mit 6-10 Klonen, die als *E. coli*-DNA tragend identifiziert wurden, im Bereich des durchschnittlich ermittelten Anteils von 6,25%. Die hohe Anzahl an Präparationswiederholungen im Falle des Klons 49E3 lässt sich sicherlich zum Teil auf die unzureichende Überprüfungsmethode zurückführen. Um mit dieser Überprüfungsmethode zuverlässigere Aussagen treffen zu können, müsste man eine deutlich größere Anzahl von Klonen sequenzieren. Das würde aber bedeuten, dass dann schon große Teile der "shotgun"-Bank sequenziert wären und die Überprüfung damit eigentlich überflüssig wäre. Bei einem mit dem letztendlich eingesetzten DNA-Präparationsprotokoll (Kapitel 2.1.1.2.4) durchschnittlich erzielten Anteil von 3,1% an Klonen, bei denen Fragmente der chromosomalen DNA der *E. coli*-Zellen in den Sequenzierungsvektor ligiert waren, erscheint dies auch sinnvoll. Eine weitere Reduzierung des Anteils an chromosomaler DNA aus *E. coli* durch weitere Verringerung von der Zellmenge zu den eingesetzten Puffervolumina scheint als wenig Erfolg versprechend.

Eine Möglichkeit bietet ein von M. S. Levy *et al.* (2000) etabliertes Verfahren, bei dem die Abtrennung chromosomaler DNA mittels Filterung durch eine Nitrozellulosemembran erfolgt. Nitrozellulose besitzt die Eigenschaft einzelsträngige DNA, RNA und Proteine zu binden. Doppelsträngige DNA hingegen wird nicht gebunden. Diese Eigenschaft machte man

sich schon zuvor zu Nutze, um einzelsträngige DNA von doppelsträngiger DNA durch Filtration und Chromatographie abzutrennen (Cohen S.N. and Miller C.A., 1965; Rush M.G. and Warner R., 1970; Wohlhieter J.A. *et al.*, 1966). Der Mechanismus der Bindung von Nukleinsäuren und Proteinen an Nitrozellulose ist komplex und noch nicht vollständig verstanden (Lui M. *et al.*, 1996). Vermutlich handelt es sich aber um hydrophobe Wechselwirkungen (Liu Y. *et al.*, 1995; Popovic and Wintzerith, 1983). Einzelsträngige DNA und RNA weist im Gegensatz zu doppelsträngiger DNA, bedingt durch frei liegende Basen, eine größere hydrophobe Oberfläche auf. In Folge dessen ist eine stärkere Wechselwirkung von einzelsträngigen Nukleinsäuren mit Nitrocellulose zu erwarten. Ferner wurde von Wohlhieter *et al.* (1966) nachgewiesen, dass die Bindung einzelsträngiger DNA an Nitrozellulose unabhängig von deren Größe ist.

Wie von M. S. Levy *et al.* (2000) gezeigt wurde, konnte der Anteil chromosomaler DNA einer versuchsweise durchgeführten Präparation des 6,9 kbp großen Plasmids pSV β von 27% auf unter 1% gesenkt werden. Dazu wurde der klare Überstand, der im Anschluss der alkalischen SDS-Lyse durch zweifache Filterung von Schwebstoffen befreit wurde, durch eine Nitrozellulosemembran gegeben. Danach erfolgte die weitere Aufreinigung durch Anionenaustauscher-Chromatographie.

Allerdings ist das bei diesen Versuchen verwendete Plasmid mit 6,9 kbp vergleichsweise klein. Bei der Herstellung von "shotgun"-Banken werden Cosmide mit einer Größe von bis zu 50 kbp, bzw. BACs und PACs mit einer Größe von bis zu 200 kbp präpariert. Sollten sich die Ergebnisse hinsichtlich der Abtrennung chromosomaler DNA übertragen lassen, so könnte sich der, im Rahmen dieser Arbeit erzielte, durchschnittliche Gehalt an chromosomaler DNA aus *E. coli* von 3.1% nochmals deutlich senken lassen.

Eine weitere Möglichkeit der DNA-Gewinnung könnte in der Verwendung des DNA Amplifikations-Kits „TempliPhi“ von Amersham Biosciences liegen (Sender A.J., 2001). Dieses Produkt ist bislang darauf ausgelegt solche Plasmidpräparationsmethoden zu ersetzen, wie sie bislang bei der Sequenzierung von "shotgun"-Banken eingesetzt werden. Der Vorteil dieser Methode liegt vor allem darin, dass für die Gewinnung von Plasmiden, die in die Sequenzierungsreaktion eingesetzt werden sollen, nur noch etwa 4 Stunden benötigt werden. Bei der normalen Plasmidpräparation werden hierfür etwa 16 Stunden benötigt. Ferner ist für die Herstellung einer ausreichenden Menge an Plasmid-DNA nur eine sehr geringe Ausgangsmenge an DNA notwendig, so dass es hinreicht, eine Zellkolonie von einer Agarplatte zu picken und in die Reaktion einzusetzen.

Dieser Prozess der DNA-Replikation wird als „Rolling Circle Amplification“ kurz „RCA“ bezeichnet. Im ersten Schritt werden die Zellen in den Reaktionspuffer des „TempliPhi“ gegeben und thermisch lysiert, und gleichzeitig wird die DNA denaturiert. Anschließend wird die RCA-Polymerase zugegeben. Wie bei der PCR bilden Primer die Startpunkte der Replikation von DNA und ein Enzym verlängert die Primer durch Einbau freier Nukleotide, um so die Ausgangssequenz zu kopieren. Allerdings wird für diesen Prozess im Unterschied zur PCR keine PCR-Maschine benötigt, sondern die Replikation erfolgt bei einer konstanten Temperatur von 30°C. Als Primer werden Gemische aus Hexameren verwendet, die jede mögliche Nukleotidsequenz repräsentieren. Deshalb binden mehrere Primer entlang der einzelsträngigen, zirkulären DNA-Matrize. Entsprechend beginnt die Replikation der DNA gleichzeitig an mehreren Stellen der DNA. Die Polymerase baut ausgehend von den Startpunkten den entsprechenden komplementären Strang auf. Trifft die Polymerase auf einen von einer anderen Polymerase aufgebauten Strang, so wird dieser von der Ausgangs-DNA abgelöst. An diese abgelöste, einzelsträngige DNA binden wieder neue Primer, und es wird wieder der komplementäre Strang aufgebaut. Da dieser Prozess immer weiter geführt wird, spricht man in diesem Zusammenhang auch von „exponentiell exponentialer Amplifikation“.

Ein weiterer Vorteil der RCA-Polymerase liegt in der sehr hohen Prozessivität dieses Enzyms, so dass es möglich zu sein scheint, zirkuläre DNA von einer und mehrerer Megabasen Länge ohne Unterbrechung replizieren zu können. Deshalb besteht unter Umständen die Möglichkeit Cosmid-, BAC- oder PAC-DNA in ausreichender Menge direkt aus einer Bakterienkolonie erhalten zu können. Durch die Verwendung von einem Hexamerengemisch als Primer wird jedoch jegliche DNA repliziert, die in den Reaktionsansatz eingesetzt wird, also auch die chromosomale DNA aus *E. coli*. Dieser Umstand spricht allerdings gegen den Einsatz dieser Methode zur DNA-Gewinnung, da man bei der Herstellung von "shotgun"-Banken gerade darauf bedacht ist, die chromosomale DNA aus *E. coli* möglichst vollständig abzutrennen. Vielleicht besteht jedoch die Möglichkeit, Primer zu planen, die unter Reaktionsbedingungen spezifisch entlang der jeweiligen Sequenz des eingesetzten Cosmids, BACs oder PACs binden. Dadurch könnte eine bevorzugte Replikation dieser DNA erzielt werden, so dass im Endeffekt der Anteil chromosomaler DNA an der Gesamtausbeute sehr gering ausfallen sollte. Somit ließe sich DNA hoher Qualität von einer großen Anzahl unterschiedlicher Klone einfach und schnell parallel herstellen. Deshalb sollte es möglich sein, dass von einer Person im gleichen Zeitintervall eine deutlich größere Anzahl an "shotgun"-Banken hergestellt werden könnte als bisher und man so dem steigenden Bedarf an "shotgun"-Banken noch besser gerecht werden könnte.

3.2 Entfernung von Klonierungsvektorsequenz tragenden Bakterienklonen aus "shotgun"-Banken

In Hybridisierungsexperimenten (Kapitel 2.2) mit einem geeigneten Sondengemisch konnte eine vollständige Entfernung von Klonierungsvektor (pBACe3.6) tragenden Bakterienklonen aus den „shotgun“-Banken der Klone 512L9 und 381K22 erzielt werden. Der Vorteil in der Verwendung eines Sondengemischs liegt im deutlich reduzierten Arbeitsaufwand. Bei den anderen in Kapitel 2.2 aufgeführten Klonierungsvektoren wurden ebenfalls keine Sequenzhomologien zu den Sequenzierungsvektoren pTZ18r und pUC18 festgestellt. Deshalb ist anzunehmen, dass dieses Verfahren auf jede "shotgun"-Bank, bei der einer dieser Vektoren zur Klonierung eingesetzt wurde, anwendbar ist. Die Ergebnisse zeigen allerdings auch, dass es zu falsch positiven Hybridisierungen kommen kann. Bei den bearbeiteten "shotgun"-Banken trat dies nur bei dem Klon 512L9 auf. In diesem Fall zeigten sich, aufgrund der ausreichenden Abdeckung durch andere DNA-Bruchstücke, keine negativen Auswirkungen auf das Gesamtsequenzierungsergebnis. Diese falsch positiven Hybridisierungen sind wahrscheinlich auf teilweise Sequenzhomologien von Klonierungsvektor und dem zu sequenzierenden DNA-Bruchstück zurückzuführen. In seltenen Fällen könnte unter Umständen eine hohe Anzahl von falsch positiven Hybridisierungen zu Lücken in der Consensussequenz führen, die trotz einer hohen Gesamtabdeckung durch Teilsequenzen auftreten.

Bei den "shotgun"-Banken 381K22 und 512L9 ist der Klonierungsvektoranteil aufgrund des Größenverhältnisses von Klonierungsvektor zu darin ligiertem DNA-Fragment mit ca. 2-3% nur gering. Die Ausfälle durch nicht angewachsene Bakterienkolonien liegen auch etwa nur bei 2%. Insgesamt ergibt sich hieraus eine verbesserte Ausnutzung der Sequenzierkapazität von 5%. Betrachtet man hingegen den Aufwand, der durch Herstellung der Filter und Durchführung der Hybridisierung entsteht, so bleibt anzumerken, dass bei solch geringen Verbesserungen der Qualität der "shotgun"-Banken, sich die Anwendung dieses Verfahren nicht lohnt. Bei unverändertem Verfahren ist die Anwendung wohl erst bei Verbesserungen von mindestens 10% sinnvoll. Wenn es jedoch gelingt, den Aufwand des Hybridisierungsverfahrens zu reduzieren, ist eine generelle Anwendung dieses Verfahrens denkbar.

Dazu könnte das gleiche Verfahren wie zur DNA-Matrizengewinnung aus Bakterienzellen (Kapitel 2.3) eingesetzt werden. Hierbei werden die Bakterienzellen direkt in den PCR-Ansatz gegeben, mittels geeigneter Primerwahl wird die jeweilige Insertion im Sequenzierungsvektor amplifiziert und im Anschluss durch Fällung aufgereinigt. Die PCR-Produkte könnten im Anschluss auf den Filter übertragen werden. Der Vorteil liegt in der Zeitersparnis, da die Vervielfältigung der DNA nicht durch Anwachsen der Bakterien-

klone über Nacht auf dem Filter erfolgt. Außerdem würde die Vorbereitung der Filter ebenfalls entfallen, die nach dem Anwachsen der Bakterienzellen notwendig ist, um diese für die Hybridisierung vorzubereiten. Dadurch besteht die Möglichkeit am selben Tag der Fertigstellung der "shotgun"-Bank Filter herzustellen und diese in die Hybridisierungsreaktion einzusetzen.

Ein zu erwartendes Problem stellt sicherlich die Übertragung einer ausreichenden Menge des PCR-Produkts auf den Filter dar, um im Vergleich zu Standard-Hybridisierungsfiltern eine vergleichbare Signalintensität zu erhalten. Der verwendete Spottingroboter (Q-Bot, Fa. Genetix) ist für die Übertragung von Bakterienzellen ausgelegt. Diese erfolgt durch Oberflächenbenetzung von Spitzen mit Bakterienkultur. Dieses nicht genau zu definierende Volumen ist ausreichend, da man die Bakterienklone auf dem Filter über Nacht wachsen lässt und so eine ausreichende DNA-Menge auf dem Filter erhält. Allerdings bietet der verwendete Spottingroboter die Möglichkeit, den Übertragungsschritt mehrfach zu wiederholen.

Daraus folgt, dass das jeweilige PCR-Produkt in einer hohen Konzentration vorliegen muss, um mit dem geringen, übertragbaren Volumen möglichst viel Produkt zu transportieren. Zusätzlich sollte das präzipitierte PCR-Produkt in einem geringen Volumen von schätzungsweise 2 µl H₂O resuspendiert werden. Das entspricht einer Konzentration von etwa 200-250 ng DNA/µl. Dadurch sollte gewährleistet sein, dass eine ausreichende Menge des PCR-Produkts auf den Filter gespottet werden kann.

In die Hybridisierungsreaktion (Kapitel 4.11) wurden durch PCR (Kapitel 4.8.2) hergestellte, DIG-11-dUTP-markierte Sonden mit einer Länge von etwa 1-1,2 kbp eingesetzt. In den Versuchen (Kapitel 2.2) wurden für den Vektor pBACe3.6 insgesamt 9 Sonden zur Identifizierung der Klonierungsvektor tragenden Bakterienklone eingesetzt. Um die Sondenherstellung umgehen zu können, bietet sich die direkte Verwendung von DIG-11-dUTP markierten Oligonukleotiden an. Für die Herstellung von "shotgun"-Banken werden Fragmentgrößen von 1 kbp bzw. 2-2,5 kbp verwendet. Ausgehend von den 1 kbp-Fragmenten, muss etwa alle 800 bp entlang der Klonierungsvektorsequenz ein Primer geplant werden, um jedes mögliche Vektorfragment mittels Hybridisierung identifizieren zu können. Ferner muss bei der Planung der Primer darauf geachtet werden, dass diese ähnliche Schmelztemperaturen aufweisen und sich, wenn möglich, nur wenig im GC-Gehalt unterscheiden. Für das Beispiel des pBACe3.6 wären dies rechnerisch mindestens 11 Primer, die als Gemisch in die Hybridisierung eingesetzt würden.

Insgesamt tragen diese beschriebenen Möglichkeiten zur generellen Nutzung dieses Verfahrens bei, allerdings verbleibt noch die arbeitsaufwendige Methode zur Entwicklung der Filter.

Für die Entwicklung wird im ersten Schritt der Antikörper Anti-DIG-AP an die DIG-11-dUTP markierten Sonden gebunden. Im zweiten Schritt wird das Substrat „AttoPhosTM“ auf den Filter gesprüht. Dieses wird durch die alkalische Phosphatase, die an den Antikörper konjugiert ist, umgesetzt. Die Detektion des Reaktionsprodukts erfolgt durch Anregung mit Breitband UV-Licht. Könnte man auf die Filterentwicklung ganz verzichten, wäre ein weiterer Schritt zum generellen Einsatz dieser Methodik erfolgt. Eine auf den ersten Blick denkbare Alternative wäre der Einsatz von radioaktiv markierten Sonden. Hier würde die arbeitsaufwendige Filterentwicklung entfallen und der Nachweis wäre sehr sensitiv. Allerdings würde hierbei ein entsprechender Sicherheitsaufwand entstehen, der keine Verringerung des Gesamtarbeitsaufwands erwarten lässt.

Alternativ wäre es denkbar in die Hybridisierung anstelle von DIG-11-dUTP markierten Sonden Primer einzusetzen, die bei der Synthese an den Enden ein- oder mehrfach mit einem Fluoreszenzfarbstoff wie z. B. Fluorescein, Tamra, Cy-3 usw. markiert werden. Laut Herstellerangaben liegt die Nachweisgrenze für Cy-dyes beim „FluorS-MultiImager“ der Fa. BIO-RAD bei 1-5 fmol. Diese Angaben sind aber auf Agarosegele bezogen, die bei der Detektion als dunkler Hintergrund erscheinen. Da die Hybridisierungsfiler durch UV-Licht ebenfalls angeregt werden, ist davon auszugehen, dass die Nachweisgrenze deutlich nach oben verschoben ist. Deshalb ist zu erwarten, dass die durch Hybridisierung gebundenen Mengen zu gering sind, um eine Signalintensität zu erhalten, die sich deutlich vom Hintergrundrauschen des Filters abhebt. Durch den Einsatz eines sogenannten „Wiener Filters“ (Wahl F. M.; 1989) könnte dieses Verfahren dennoch einsetzbar sein. Beim „Wiener Filter“ handelt es sich um einen Algorithmus, der eine Entfernung von Hintergrundrauschen ermöglicht. Dadurch würde eine Entwicklung der Filter entfallen, und durch die Ungefährlichkeit dieser Farbstoffe könnte auch die Problematik vermieden werden, wie sie im Umgang mit Radioaktivität entsteht.

Aufgrund dieser Ausführungen wäre auch ein Verfahren denkbar, bei dem man ganz auf die Verwendung von Filtern für die Hybridisierung verzichten könnte. Die Filter dienen zur Immobilisierung der DNA, an die entsprechende Sonden hybridisieren können. Im oberen Teil wurde diskutiert, PCR-Produkte auf den Filter zu übertragen und an der Oberfläche des Filters zu immobilisieren. Diese Immobilisierung könnte man auch erreichen, wenn man in die PCR-Reaktion Primer einsetzt, die am 5'-Ende mit Biotin markiert sind. Diese würden mit dem PCR-Produkt an die Steptavedinoberfläche von "Dynabeads" binden. Durch die magnetischen Eigenschaften der "Dynabeads" kann das PCR-Produkt durch Anlegen eines Magnetfelds bei Waschschritten zur Reinigung des Produkts in den PCR-Gefäßen zurückgehalten

werden. Deshalb wäre es denkbar, das oben beschriebene, fluoreszenzmarkierte Primergemisch in 1x PCR-Puffer gelöst auf das aufgereinigte PCR-Produkt zu geben. Nach der Denaturierung der PCR-Produkte bei 95°C im PCR-Gerät könnte eine spezifische Bindung der Primer an PCR-Produkte mit Klonierungsvektorsequenz durch Wahl der Temperatur nahe der Schmelztemperatur der Primer gewährleistet werden. Durch ein- bzw. mehrmaliges Waschen mit 1x PCR-Puffer und eventuellem Aufheizen der Proben bis zur Schmelztemperatur des Primergemischs würden überschüssige bzw. nicht gebundene Primer entfernt werden. Dieses Verfahren hätte den Vorteil, dass die Hybridisierung im Gegensatz zur Verwendung von Hybridisierungsfiltren nicht über Nacht, sondern wie beim Annealing-Schritt eines PCR-Zyklus innerhalb weniger Minuten erfolgt. Außerdem ist ein Mengenproblem, wie bei der Übertragung des PCR-Produkts auf den Hybridisierungsfiltren, durch die Möglichkeit das gesamte PCR-Produkt in die Hybridisierung einsetzen zu können, nicht zu erwarten.

Um bei der Detektion störende Einflüsse durch Fluoreszenzsignale benachbarter Gefäße der PCR-Platte auszuschließen, sollten anstelle der normalerweise lichtdurchlässigen Platten, schwarze PCR-Platten verwendet werden. Die Detektion könnte ebenfalls mit dem „FluorS-MultiImager“ der Fa. BIO-RAD erfolgen. Eine bessere Variante wäre aber die Detektion mit einem Mikrotiterplattenlesegerät zur Auswertung fluoreszenzbasierter Anwendungen, wie z. B. dem Gerät „GENIOS“ der Fa. Tecan, da dieses Gerät einen empfindlicheren Messbereich aufweist und sich außerdem für eine Einbindung in eine automatisierte Umgebung eignet. Ferner kann dieses Gerät, durch den Einsatz von entsprechenden Ausschlussfiltren, sehr variabel auf die Absorptions- bzw. Emissionswellenlängen unterschiedlicher Fluoreszenzfarbstoffe eingestellt werden. Laut Herstellerangaben liegt die Nachweisgrenze für Fluorescein bei 8 fmol in 200 µl Lösung pro Gefäß. Bei der Amplifikation eines etwa 1,2 kbp großen DNA-Inserts mittels PCR direkt aus Bakterienzellen erhält man ca. 750 ng Produkt (Kapitel 2.2). Bei dieser Fragmentgröße entsprechen 750 ng DNA ca. 1 pmol. Das bedeutet, dass nur ein Bruchteil von vielleicht 50-100 fmol des PCR-Produkts in die Hybridisierung eingesetzt werden muss, um eine sichere Detektion zu gewährleisten.

Bei den vorgestellten alternativen Verfahren wird jeweils nur eine geringe Menge des PCR-Produkts für die Hybridisierungen verwendet. Die Restmenge an PCR-Produkt sollte deshalb ausreichend sein, um in eine nachfolgende Sequenzierungsreaktion eingesetzt werden zu können. Nachdem die Hybridisierungsergebnisse ausgewertet sind, müsste allerdings zuvor eine Umorganisation der PCR-Produkte in eine neue Mikrotiterplatte erfolgen, um eine Entfernung der Klonierungsvektorfragmente und der Positionen, bei denen aufgrund der nicht angewachsenen Bakterienkolonien kein PCR-Produkt erhalten wurde, zu gewährleisten. Diese

Umsortierung könnte durch einen dafür geeigneten Pipetierroboter erfolgen. Parallel würde die Umsortierung der Kulturplatten mit dem Picking- und Spottingroboter „Q-bot“ der Fa. Genetix erfolgen. Deshalb müsste die Möglichkeit gegeben sein, durch entsprechende Einstellungen der unterschiedlichen Systeme jeweils das gleiche Umsortierungsschema verwenden zu können. Nur so ist gewährleistet, dass bei eventuell notwendigen Kontrollsequenzierungen der richtige Bakterienklon verwendet wird.

Die Realisierbarkeit der vorgestellten Verbesserungen muss in entsprechenden Experimenten überprüft werden. Sollten sich nicht alle Teilschritte als akzeptabel herausstellen, so muss überprüft werden, ob sich das entsprechend realisierte Verfahren für die generelle Anwendung eignet. Sollte sich das Verfahren in der letztendlich ausgearbeiteten Form realisieren lassen, wäre eine entsprechende Anwendung sehr sinnvoll, um so die Qualität der „shotgun“-Banken steigern zu können. Ferner wäre die Möglichkeit gegeben, dieses Verfahren in eine zumindest teilweise automatisierte Umgebung zu integrieren.

Eine „shotgun“-Bank weist durchschnittlich einen Anteil von ca. 10% Klonierungsvektor tragenden Bakterienklonen (Kapitel 2.2) auf. Durch Entfernung dieser Klone aus der Bank ergeben sich im Bereich der Genomsequenzierung mehrere Vorteile. Die vorhandenen Sequenzierkapazitäten können ökonomischer ausgenutzt werden. Die Durchführung von Sequenzierprojekten mit diesen verbesserten „shotgun“-Banken kann dadurch 10% schneller erfolgen. Ferner trägt dieses Verfahren zur Reduzierung an den Gesamtkosten in der Genomsequenzierung bei.

3.3 Sequenzierung von PCR-Produkten

Bisher wurden schon verschiedene Ansätze der „direkten Sequenzierung“ publiziert. Diese eignen sich aber aus unterschiedlichen Gründen nicht oder nur bedingt für den Einsatz in der Mega-Sequenzierung. Bei einem dieser Beispiele wird die hohe Sensitivität der Sequenzierung mit Radio-Nukleotiden ausgenutzt (Krishnan B. R. *et al.*, 1991). Das verwendete Protokoll lässt sich aber nicht auf die heutige Sequenzier Technologie übertragen. Oder es wird auf eine Aufreinigung des erhaltenen PCR-Produkts verzichtet (Kilger C. and Schmid K., 1994). Die verwendeten PCR-Primer und Nukleotide werden mit dem PCR-Produkt in die Sequenzierungsreaktion geschleppt und führen zu verkürzten Leseweiten und schlechter Qualität. Bei einem weiteren Protokoll erfolgt die Aufreinigung der PCR-Produkte mit einem kommerziell erhältlichen Kit (Moslemi A.-R. *et al.*, 1997). Solche Kits sind ebenfalls auf ein 96er-Format beschränkt und führen zu den gleichen Problemen wie Plasmid-Präparationskits. Es wurden in den Publikationen aber auch zwei Protokolle gefunden, bei denen eine Aufreinigung des

PCR-Produkts mittels Fällung erfolgt. Im ersten Protokoll erfolgt die Fällung ausgehend von einem 100 µl PCR-Ansatz durch Zugabe von Ethanol und 3 M NaAc in einem Endvolumen von ca. 300 µl (Ohno K. *et al.*, 1991). Generell wäre dieses Protokoll durch Verringerung des PCR-Ansatzes auf 15 µl auch für die Durchführung in einer PCR-Platte im 384er-Format (je 50 µl Volumen/Gefäß) geeignet. Unter Verwendung dieses Protokolls beschränken sich laut Angaben die erzielbaren Leseweiten auf 300-400 bp. Im Vergleich wurden mit dem im Rahmen dieser Arbeit etablierten Protokoll problemlos Leseweiten von 500-650 bp erreicht. Im zweiten Protokoll wird die Fällung ausgehend von einem 20 µl PCR-Ansatz durch Zugabe von H₂O, Isopropanol und 5 M NaCl in einem Endvolumen von 200 µl durchgeführt (Hodall E. *et al.*, 1999). Bei diesem Fällungsverfahren besteht nicht die Möglichkeit, die Volumina zu reduzieren und so an eine Durchführung in einer PCR-Platte im 384er-Format anzupassen, da das Volumen des PCR-Ansatzes zu gering wäre. Aus diesem Grund bleibt dieses Verfahren auf eine Durchführung in PCR-Platten im 96er-Format (je 200 µl Volumen/Gefäß) beschränkt. Ferner wurden in dem Artikel keine Angaben zu den maximal erzielbaren Leseweiten gemacht.

Im Gegensatz zu diesen Ansätzen konnte ein Protokoll (Kapitel 4.8.1) für die „direkte Sequenzierung“ erarbeitet werden, das sich uneingeschränkt in der Mega-Sequenzierung einsetzen lässt. Im Vergleich zur Standard-Methode (Sequenzierung präparierter Plasmid-DNA) werden Sequenziererergebnisse von gleicher Qualität erzielt. Im Gegensatz hierzu ist das vorgestellte Verfahren nicht auf das 96er-Format beschränkt, sondern ist vor allem auf eine Durchführung im 384er-Format ausgelegt. Eine Übertragung auf ein 1532er-Format wäre unter entsprechenden Modifikationen des Protokolls ebenfalls denkbar. Allerdings gibt es bislang noch keine PCR-Geräte, die auf dieses Format ausgelegt sind. Durch die Etablierung dieses Protokolls ist es gelungen einerseits durch Verringerung des Arbeitsaufwands und andererseits durch die Variationsmöglichkeiten im verwendeten Format eine deutliche Erhöhung im Probendurchsatz zu erzielen. Zusätzlich bietet die DNA-Gewinnung für den Einsatz in die Sequenzierung mittels PCR im Vergleich zu kommerziell erhältlichen Plasmid-Präparationskits eine Kosteneinsparung von mindestens 80%. Ferner eignet sich dieses Verfahren auf relativ einfache Weise für die Durchführung in einer automatisierten Umgebung. Dies legt nahe, die kosten- und arbeitsaufwendige Plasmid-Präparation durch das beschriebene Verfahren zu ersetzen. Die technische Entwicklung im Bereich der Sequenzierautomaten, speziell durch die Einführung der Kapillarelektrophoreseautomaten, erlaubt immer höheren Sequenzierdurchsatz. Bei der Verwendung von handelsüblichen Plasmid-Präparationskits, die auf eine Durchführung im 96er-Format beschränkt

sind, ist deshalb im Bereich der Mega-Sequenzierung ein Engpass abzusehen, der unter Beibehalt der Methode nur durch die Anschaffung einer entsprechenden Anzahl kostspieliger Präparationsroboter zu beheben ist. Durch „direkte Sequenzierung“ ist es möglich, die Sequenzierkapazitäten heutiger Sequenzierautomaten zu nutzen, ohne den immensen technischen Aufwand betreiben zu müssen, wie er bei der Verwendung der Standardmethode nötig wäre.

Eine alternative Möglichkeit der DNA-Gewinnung direkt aus Bakterienklonen bietet der kürzlich auf den Markt gekommene, in Kapitel 3.1 beschriebene, DNA-Amplifikations-Kit „TempliPhi“ von Amersham Biosciences (Sender A.J., 2001). Mit diesem Verfahren ist es möglich, in einer Reaktion, die als „Rolling Circle Amplification“ beschrieben wird, direkt aus Bakterienzellen große Mengen DNA zu erhalten, die anschließend ohne Aufreinigung in die Sequenzierreaktion eingesetzt werden kann. Dieses Verfahren ist mit geringem Arbeitsaufwand verbunden und dürfte sich ebenfalls für eine Automation eignen.

In Kapitel 3.2 sind verschiedene Möglichkeiten beschrieben, um Klonierungsvektor tragende Bakterienklone mittels Hybridisierung aus „shotgun“-Banken zu entfernen. Da sich beim „TempliPhi“ in der Enzymlösung auch das Hexamerengemisch als Primer befindet, besteht nicht die Möglichkeit, diese durch ein Hexamerengemisch zu ersetzen, das am 5'-Ende mit Biotin markiert ist. Um Hybridisierungsexperimente durchführen zu können, muss deshalb die so amplifizierte DNA auf Filter übertragen werden. Dadurch ist die Entfernung Klonierungsvektor tragender Bakterienklone aus einer „shotgun“-Bank mit einem entsprechenden Mehraufwand verbunden. Aus diesem Grund ist ein vollautomatischer Prozess, der die DNA-Gewinnung, die Entfernung Klonierungsvektor tragender Klone und die anschließende Durchführung der Sequenzierreaktion umfasst, sicherlich nur durch Trennung der Komponenten des TempliPhi-Kits realisierbar.

Im Vergleich hierzu ist ein solcher Prozess mit der beschriebenen DNA-Gewinnung durch PCR-Amplifikation unter Verwendung von „Dynabeads“ und Primern, die am 5'-Ende mit Biotin markiert sind, wohl einfacher zu realisieren. Für die Verwendung von „Dynabeads“ sind automatische Vorrichtungen erhältlich, die sich in eine automatisierte Umgebung integrieren lassen. Der ohnehin schwer automatisierbare Fällungsschritt zur Aufreinigung des PCR-Produkts würde dadurch entfallen. Unter diesen Gesichtspunkten ließe sich der zusätzliche Kostenaufwand für die „Dynabeads“ oder ein ähnliches Produkt durchaus vertreten.

3.4 Sequenzierung und Annotation

Die Sequenzierung der Klone 98L15 und 49E3 von Chromosom 21 (Kapitel 2.4) gestaltete sich unterschiedlich schwierig. Der Klon 98L15 wies nach der Assemblierung der „shotgun“-Sequenzen eine ähnliche Anzahl von größeren Contigs auf wie der Klon 49E3. Die Lücken zwischen den einzelnen Contigs waren aber wesentlich kleiner als beim Klon 49E3 und ließen sich relativ einfach durch „primer walking“ schließen. Die Situation, wie sie beim Klon 98L15 vorgefunden wurde, stellt den Normalfall dar. In ähnlicher Form ließen sich auch die Sequenzierungen der anderen Klone durchführen, von denen im Rahmen dieser Arbeit „shotgun“-Banken hergestellt wurden (Kapitel 2.1). Es gab zwar Schwankungen in der Anzahl der Contigs, die nach der Assemblierung der „shotgun“-Sequenzen erhalten wurden, die Lücken zwischen diesen Contigs ließen sich aber allgemein relativ einfach durch „primer walking“ schließen. Bei der Sequenzierung von Klonen, die nur ein relativ kleines Fragment trugen (Cosmide), kam es teilweise sogar vor, dass die lückenfreie Consensussequenz schon nach der Assemblierung der „shotgun“-Sequenzen erhalten wurde.

Die Probleme, die bei der Sequenzierung des Klons 49E3 auftraten, sind somit als Sonderfall zu betrachten. Bis auf die ca. 8,5 kbp große Lücke zwischen den Contigs E und F (Kapitel 2.4) sind die Lücken zwischen den anderen Contigs mit 120 bp bis 3 kbp als normal zu betrachten. Die außergewöhnlich große Lücke zwischen den Contigs E und F lässt sich vermutlich auf die Fragmentierung der DNA zurückführen. Wahrscheinlich fragmentierte die DNA in diesem Bereich besonders stark, so dass dieser Bereich nach der Längenselektion der DNA-Fragmente unterrepräsentiert vorlag. Hinzu kam noch, dass ein Schließen der Lücken durch „primer walking“ in diesem Fall nicht möglich war, da die erhaltenen Sequenzierungsergebnisse aufgrund der schlechten Qualität unbrauchbar waren. Dies lässt sich wahrscheinlich auf die verwendete Sequenzierreaktion zurückführen, die standardmäßig nach dem ABI-Protokoll (Kapitel 4.8.3) durchgeführt wurde, weil hier keine teuren fluoreszenzmarkierten Primer notwendig sind. Bei der Wiederholung einiger Sequenzierungen nach dem LI-COR-Protokoll (Kapitel 4.8.3) mit gleichen Primern, die aber entsprechend fluoreszenzmarkiert waren, konnten einwandfreie Sequenzierungsergebnisse erhalten werden. Zusätzlich besteht aber auch die Möglichkeit, dass einige Primer in den Reaktionen nach dem ABI-Protokoll schlecht oder mehrfach gebunden haben und deshalb schlechte Sequenzierungsergebnisse erhalten wurden. Trotz dieser Probleme war es möglich, die Sequenz dieses Klons durch Sequenzierung auf PCR-Produkten bzw. auf klonierten Restriktionsfragmenten vollständig zu erhalten. Diese Vorgehensweise wurde durch die große Überlappung mit dem Klon R73i15 (RIKEN) begünstigt. Dadurch war es möglich, die Reihenfolge der Contigs sicher festzulegen und die Größe

der Sequenzlücken zwischen diesen zu ermitteln. Ferner war es dadurch möglich, unter Kenntnis der Sequenz innerhalb der Lücken, Primer für die Herstellung von PCR-Produkten zu planen, um auf ihnen die Sequenzierung durchführen zu können.

Die angewandte Methode lässt sich aber teilweise auf die Sequenzierung von Klonen übertragen, bei denen keine Informationen hinsichtlich der Lücken vorliegen und eine Sequenzierung durch „primer walking“ aus den oben genannten Gründen nicht möglich ist. Nach der Planung von Primern an den Enden der Contigs kann jede mögliche Primerkombination in eine PCR-Reaktion eingesetzt werden. Wird mit einer Primerkombination ein Amplifikationsprodukt erhalten, so lassen sich die entsprechenden Contigs einander zuordnen. Anschließend kann das erhaltene Amplifikationsprodukt in die Sequenzierreaktion eingesetzt werden. Andererseits kann die DNA eines Klons mit unterschiedlichen Restriktionsenzymen geschnitten und dann gelelektrophoretisch aufgetrennt werden. Nach der Übertragung der DNA aus dem Gel auf eine Nitrocellulose-Membran mittels Western-Blot können, nach der Hybridisierung mit geeigneten Sonden, Restriktionsfragmente identifiziert werden, die Sequenzlücken abdecken. Diese Fragmente können dann isoliert und kloniert werden. Danach kann die Sequenzierung mit diesen Klonen durchgeführt werden.

Nach der vollständigen Sequenzierung der beiden Klone wurden die Sequenzen im Überlappungsbereich benachbarter Klone verglichen. Wie gezeigt (Kapitel 2.4) stimmten die Sequenzen der Klone 98L15 und 49E3 innerhalb des Überlappungsbereichs vollständig überein. Auch die Sequenzen der Klone 98L15 und 90B5 (IMB Jena) stimmten hier mit Ausnahme einer Deletion im Klon 90B5 (IMB Jena) überein. Da die Klone aus der selben Bibliothek stammen, die aus der DNA eines Individuums erzeugt wurde, sind keine Sequenzunterschiede zu erwarten. Die gefundene Deletion ist deshalb vermutlich auf einen Fehler während der DNA-Replikation innerhalb der Bakterienzellen zurückzuführen.

Die benachbarten Klone 49E3 und R73i15 (RIKEN Japan) stammen aus unterschiedlichen Bibliotheken. Deshalb sind hier im Überlappungsbereich sehr wohl Sequenzunterschiede zu erwarten, die auf die unterschiedliche Entwicklung zweier Individuen zurückzuführen sind. Innerhalb des Überlappungsbereichs sind insgesamt 31 Deletionen, 208 Einzel- und 3 Mehrfachbasenaustausche gefunden worden. Da bislang erst die Sequenz des Klons 49E3 auf Editierungsfehler geprüft wurde, eine Überprüfung des Klons R73i15 von Seiten des „RIKEN Genomic Sciences Center“ (Japan) aber noch aussteht, kann sich die gefundene Anzahl der Sequenzunterschiede noch deutlich verringern. Deshalb wird auf eine statistische Betrachtung der Häufigkeit der gefunden Sequenzunterschiede vorerst verzichtet. Allerdings fällt bei den 31 Deletionen vor allem die Deletion 4 (Kapitel 7.3.2.1) ins Auge. Hier findet

sich auf dem Klon 49E3 eine 152 bp lange Sequenz, an deren Stelle auf dem Klon R73i15 (RIKEN Japan) eine 1452 bp lange Sequenz zu finden ist. Diese Sequenzen zeigen keine Homologien. Es konnten in diesem Bereich auch keine Transposons gefunden werden, die einen Hinweis liefern, dass diese Sequenzunterschiede durch die Einwirkung eines Virus hervorgerufen wurden. Für das Schließen der Sequenzlücken beim Klon 49E3 wurden, wie schon erwähnt, unter anderem Restriktionsfragmente verwendet. Hierbei wurde das zu erwartende Verdauemuster anhand der vorliegenden Consensussequenz theoretisch berechnet, um die benötigten Fragmente identifizieren zu können. Da die theoretischen Berechnungen mit den experimentellen Ergebnissen übereinstimmen, ist eine fehlerhafte Assemblierung zumindest für den Klon 49E3 ausgeschlossen. Es erscheint deshalb als sinnvoll aus der jeweiligen Bibliothek weitere Klone herauszusuchen, die diesen Bereich abdecken und diesen nochmals zu sequenzieren.

Durch Annotation der Sequenzen der Klone 98L15 und 49E3 wurden mögliche Gene vorhergesagt (Kapitel 2.4). Im Fall des Klons 98L15 wurde die Annotation zweimal durchgeführt. Das erste mal im Juni 1999 nach vollständiger Sequenzierung des Klons. Mittels dieser Daten konnten zwei Gene gefunden werden. Das vollständige „RNA binding motif protein 11“- Gen und das unvollständige Triglycerid Lipase-Gen (EC 3.1.1.3). Von diesem Gen konnten zu diesem Zeitpunkt 3 Exons gefunden werden. Die zweite Annotation erfolgte im Dezember 2001 zusammen mit dem Klon 49E3. Durch die aktuellere EST-Datenbank war es möglich, das zuvor auf dem Klon gefundene unvollständige Triglycerid Lipase-Gen (EC 3.1.1.3) um ein weiteres Exon auf 4 Exons zu verlängern. Auf dem Klon 49E3 wurde bislang noch kein Gen gefunden. Durch weiter vervollständigte EST-Datenbanken, der Möglichkeit mehr ähnliche Sequenzen aus anderen Genomprojekten heranziehen zu können und durch den Einsatz von Exon-Vorhersageprogrammen mit verbesserten Algorithmen sollten sich sicherlich noch weitere Gene finden lassen bzw. unvollständige Gene vervollständigt werden können.

3.5 Einsatzmöglichkeiten des Picking-Tools

Mit der Entwicklung des Picking-Tools (Kapitel 2.5) ist es gelungen, eine einfache und kostengünstige Vorrichtung zu konstruieren, mit der es möglich ist, sehr schnell Bakterienkulturen von Agarplatten in Mikrotiterplatten im 96er- oder 384er-Format zu übertragen. Mit dieser Vorrichtung kann man vier Mikrotiterplatten im 384er-Format innerhalb einer Stunde beimpfen. Wie schon erwähnt, stößt man damit in Bereiche von Pickgeschwindigkeiten vor, die bislang ausschließlich Robotersystemen vorbehalten waren.

Allerdings ist das Picking-Tool nicht generell als Alternative zu Robotersystemen gedacht. Schon deshalb nicht, da Roboter oftmals nicht nur auf das Picken von Bakterienkolonien beschränkt sind, sondern auch für das Spotten von Hybridisierungsfiltern, die Umorganisation von Bakterienklonen innerhalb von Mikrotiterplatten und das Kopieren solcher Platten ausgelegt sind. Auch für den generellen Einsatz in Sequenzierlaboren, die ausschließlich die Picking-Option eines solchen Robotersystems nutzen würden, aber einen so großen Bedarf an "shotgun"-Banken aufweisen, dass ein solches System ganztägig eingesetzt würde, ist das Picking-Tool nicht ausgelegt. Einerseits liegt die Übertragungsgeschwindigkeit der meisten Systeme höher. Andererseits würde der ganztägige und ununterbrochene Einsatz des Picking-Tools bei der ausführenden Person sicherlich zu Ermüdungserscheinungen führen, die zwar nicht unbedingt zu Fehlern, wohl aber zu einer verringerten Pickgeschwindigkeit führen würden. Sehr wohl ist der Einsatz des Picking-Tools auch in oben beschriebenen Sequenzierlaboren denkbar um Ausfälle durch Reparaturen am Roboter annähernd kompensieren zu können. Bei der Firma Lion-Bioscience in Heidelberg wurde das Picking-Tool der ersten Generation auf mögliche Einsatzzwecke getestet. Da die Firma eine relativ große Sequenzierabteilung aufweist und im Besitz eines Picking- und Spottingroboters ist, stand ein genereller Einsatz des Tools von vornherein außer Frage. Es hat sich aber gezeigt, dass die Verwendung des Picking-Tools immer dann sinnvoll und zeitsparend war, wenn es darum ging eine relativ geringe Anzahl von Bakterienklonen, etwa zum Animpfen von 1-2 Mikrotiterplatten im 384er Format, picken zu müssen. In solchen Fällen war es zu zeitaufwendig und zu umständlich den Roboter für das Picken einzurichten. Auch wenn eine "shotgun"-Bank mittels Roboter gepickt wurde und die Anzahl an Bakterienkolonien auf der Agarplatte nicht ausreichten, um die letzte Mikrotiterplatte der Bank vollständig anzuimpfen, hat es sich als günstiger erwiesen, das Picking-Tool zum Vervollständigen dieser Mikrotiterplatte einzusetzen. Denn es war zeitaufwendiger, eine weitere Agarschale mit Bakterienklonen in den Roboter zu stellen, diese mittels der Kamera aufzunehmen und von dem Programm auswerten zu lassen und anschließend das Picken durchzuführen.

Dabei hat das Picking-Tool durch die erzielbare Geschwindigkeit, mit der manuell Bakterienkolonien aufgenommen und in Mikrotiterplatten übertragen werden können, überzeugt. Vor allem aber auch dadurch, dass sich diese manuelle Übertragung absolut fehlerfrei hinsichtlich ausgelassener oder doppelt beimpfter Positionen einer Mikrotiterplatte, wie es bei der Verwendung von Zahnstochern häufiger vorkommt, ausführen ließ.

Dieses Resultat zeigt, dass auch in großen Sequenzierlaboren der parallele Einsatz des Picking-Tools neben einem Robotersystem durchaus sinnvoll ist, wenn nur eine geringere Anzahl von Bakterienklonen zu picken ist.

Für kleinere Sequenzierlabore bildet das Picking-Tool eine echte Alternative zu kostenintensiven Robotersystemen. Vor allem dann, wenn der Bedarf an "shotgun"-Banken zu groß ist, um ihn durch Picken mit Zahnstochern decken zu können, aber ansonsten zu klein ist, als dass sich die Anschaffung eines entsprechenden Roboters auszahlen würde.

Eine weitere Einsatzmöglichkeit bietet sich in Laboren, die "shotgun"-Banken herstellen, um anschließend daraus Hybridisierungsfilter für Kartierungen herstellen zu können. Durch diese Kartierung werden Überlappungsbereiche einzelner Klone ermittelt. Durch diese Sortierung ist es möglich, mit einer minimalen Anzahl von Klonen die Ausgangssequenz abdecken zu können. Da die Kartierung solcher "shotgun"-Banken den Hauptarbeitsprozess darstellt, wird entsprechend eine geringe Anzahl an "shotgun"-Banken benötigt, so dass sich eine Anschaffung eines Pickingroboters meist nicht rentiert. In solchen Laboren erfolgt die Übertragung der Bakterienkolonien von Agarplatten in Mikrotiterplatten meist mit sterilen Zahnstochern. Im Anschluss erfolgt die Übertragung der Bakterienkolonien aus den Mikrotiterplatten auf die Filter meist mittels eines Spottingroboters wie den „BioGrid“ der Fa. Biorobotics, der ausschließlich diese Funktion bietet und deshalb relativ kostengünstig ist. Gerade in solchen Laboren bietet sich der Einsatz des Picking-Tools zur Übertragung der Bakterienkolonien in Mikrotiterplatten an.

Daraus ergibt sich ein breites Einsatzgebiet des Picking-Tools, wenn es ausschließlich darum geht, Bakterienkolonien von Agarplatten in Mikrotiterplatten zu übertragen. Dabei schließt das Picking-Tool von seiner Konzeption die Lücke zwischen der langsamen, fehlerbehafteten Bakterienübertragung mit sterilen Zahnstochern und sehr schnellen aber teuren Robotersystemen.

4 Material und Methoden

4.1 Geräte

Tischzentrifugen:

Minifuge Hereaus-Christ, Biofuge pico

Eppendorf, Centrifuge 5417C

Kühlzentrifuge

Rotoren

Sorvall RC5B

SS34, GS3 und GSA

Ultrazentrifuge

Rotor

Sorval OTD Combi

Ti 70

Mikrotiterplatten-Zentrifuge:

Jouan CR 322

Gelelektrophoresekammern:

BIO-RAD Sub-Cell GT 25cm

BIOMETRA Agagel Maxi 20cm

Netzgeräte:

BIO-RAD, POWER-PAC 3000

Pharmacia ECPS 3000/150

Geldokumentationssystem

BIO-RAD MultiImager

Elektroporation

BIO-RAD Genepulser mit Capacitance Extender und Pulse controller

Picking- und Spottingroboter

GENETIX Q-bot

Mikrotiterplatten-Befüllsystem

GENETIX Q-Fill

Sequenzier:

Perkin Elmer ABI PRISM™ 377

LI-COR, dna sequencer LONG READIR 4200

Zellkultur-Inkubatoren:

Adolf-Kühner AG LAB-shaker und LAB-Therm

Infors UNITRON

PCR-Maschinen

MJ Research PTC-200

MWG-Biotech Primus 96^{plus} mit 96er bzw. 384er PCR-Einsatz

Thermomixer

Eppendorf Thermomixer 5436

Heizblock

Techne Dri Block DB3

pH-Meter

WTW pH526

Vortex

IKA MS1 Minishaker mit zusätzlicher Halterung für Mikrotiterplatten

Sterilbanken

Heraeus Lamin Air HLB 2448

Heraeus Hera safe HS 18/2

4.2 Chemikalien und Enzyme

AMERSHAM GmbH, Braunschweig:

Thermo Sequenase fluorescent labelled primer cycle sequencing kit with 7-deaza-dGTP, Pimercycle (‘Formamidhaltiger Gelbeladungspuffer)

BIORAD GmbH, München:

30% ige Acrylamid-Lösung, Mischungsverhältnis Acrylamid/Bisacrylamid beträgt 29:1

BIOZYM GmbH, Olendorf:

Plasmid-SafeTM, SeaKem LE Agarose, Hydrolink

BOEHRINGER Mannheim GmbH:

EcoR I, Sma I, Pronase, T4-DNA-Ligase, alkalische Phosphatase (CIP), DIG-dUTP alkalistable, Anti-Digoxigenin-AP, AttophosTM Substrat Set, Herring Sperm DNA, Blocking agent, Molecular Weight Marker IV

DIFCO LABORATORIES, Detroit USA:

Bacto-Yeast-Extract, Bacto-Tryptone, Bacto-Agar

FLUKA Chemica, Seelze:

Natrium-Dodecsulfat, Kaliumacetat, Kaliumiodid, Kaliumchlorid, β -Mercaptoethanol, Magnesiumchlorid, Magnesiumacetat, Diethanolamin, Xylencyanol

Gibco GmbH:

1 kb-DNA-Leiter

MERK GmbH, Darmstadt:

Borsäure, Glucose, Glycerin, konz. Essigsäure, konz. Salzsäure, Natriumacetat, Natriumchlorid, Natriumdihydrogenphosphat, di-Natriumhydrogenphosphat, Natriumhydroxid, Phosphorsäure

New England Biolabs, Frankfurt am Main:

Vent DNA Polymerase, T4-DNA-Polymerase, Klenow-Fragment

PERKIN ELMER GmbH,

Ready Reaction DyeDeoxy Terminator AmpliTaqFS Kit

PHARMACIA GmbH, Freiburg:

Ultrapure dNTP-Set

QIAGEN GmbH, Hilden:

Taq-Polymerase

Riedel-de Haen Ag, Seelze :

Ammoniumsulfat

ROTH, Karlsruhe:

X-Gal, IPTG

SERVA GmbH, Heidelberg:

APS, TEMED, MOPS, Bromphenolblau, Sarcosyl NL-30, Triton X-100, Tris-acetat, Tris-Base, Tris-HCl

SIGMA GmbH, Deisenhofen:

DTT, Ampicillin, Kanamycin, Chloramphenicol, Ethidiumbromid, RNase A

USB, Bad Homburg:

Harnstoff „ultragrade“ zur Sequenzierung

4.3 Materialien

4.3.1 Bakterienstämme

XL1-Blue : F' ::Tn10(Tet^R) *proA* + *B* + *lacI*_q (*lacZ*) *M15/recA1 endA1 gyrA96* (NaI^r) *thi hsdR17*(^{rK} - ^{mK} +) *supE44 relA1 lac* (Bullock et al., 1987).

DH10B: F' *mcrA*; Δ(*mrr-hsdRMS-mcrCB*); Φ80*lacZ*Δ*M15*; Δ*lacX74*; *deoR*; *recA1*; *AraD139*; Δ(*ara-leu*)7697; *galU*; *galK*; λ'*rpsL*; *nupG*

Top10F': F'(*lacI*_q, Tn10(*tet*^r), *mcrA*, D(*mrr-hsdRMS-mcrCB*), f80*lac*□Δ*M15*, Δ*lacX74*, *deoR*, *recA1*, *araD139*, D(*ara-leu*)7697, *galU*, *galK*; (Invitrogen)

4.3.2 Plasmide

pTZ-18R (PHARMACIA GmbH, Freiburg)
pUC 18 (Vieira & Messing, 1982).

4.3.3 Nährmedien und Antibiotika

LB	5g/L Bacto-Yeast-Extrakt, 10g/L Bacto-Trypton, 5g/L NaCl
2YT	10g/L Bacto-Yeast-Extrakt, 16g/L Bacto-Trypton, 5g/L NaCl
SOC	5g Bacto-Yeast-Extrakt, 20g Bacto-Trypton, 0,5g NaCl in 950 ml bidest. H ₂ O gelöst und autoklaviert, dann Zugabe von 10 ml 250 mM KCl (steril), 5 ml 2 M MgCl ₂ (steril) und 20 ml 1 M Glucose (sterilfiltriert), auf pH 7,0 mit 1 M NaOH eingestellt

Ampicilin	100 mg / ml in H ₂ O
Kanamycin	50 mg / ml in H ₂ O
Chloramphenicol	12,5 mg / ml in Ethanol

4.3.4 Oligonukleotide

Standartsequenzierprimer:

ABI-Primer:

UP0	5'-CTG CAG GTC GAC TCT AGA GGA TC-3'
RP0	5'-CTA TAG GGA ATT CGA GCT CGG TAC-3'
M13/pUC (-20)	5'-GTA AAA CGA CGG CCA GT-3'
M13/pUC (-26)	5'-CAG GAA ACA GCT ATG AC-3'

LI-COR-Primer:

Die LI-COR-Primer sind am 5'-Ende mit den Farbstoffen IRD700 oder IRD800 fluorensenzmarkiert.

UNI 700	5'-AGG GTT TTC CCA GTC ACG ACG TT-3'
REV 800	5'-GAG CGG ATA ACA ATT TCA CAC AGG-3'

Bei der Planung aller weiteren Primer (Kapitel 7.2) wurden nach Möglichkeit folgende Kriterien zugrundegelegt:

- Am 5'-Ende GC-reich beginnend und am 3'-Ende AT-reich enden
- Die Länge der Primer sollte zwischen 21 und 29 Basen liegen
- Der GC-Gehalt sollte 30-50% betragen
- Die Berechnung der Schmelztemperatur erfolgte mit dem „Calculator“ der Firma Interactiva (www.interactiva.de) und lag zwischen 53-59°C.

Primer für PCR-Produktgewinnung für die Sequenzierung:

Die Primer wurden für den Sequenzierungsvektor pUC 18 geplant:

pUC 18.for	5'-GAT GTG CTG CAA GGC GAT-3'
pUC 18.rev	5'-GGC ACC CCA GGC TTT ACA-3'

Primer für Hybridisierungsexperimente:

Primer zur Herstellung von Sonden zur Entfernung des Klonierungsvektors pBACe3.6 aus „shotgun“-Banken. Insgesamt wurden 9 Sonden hergestellt, die den Vektor vollständig abdecken.

Sonde 1:

BAC36-1F	5'-GCC AGG CCT ACC CAC TAG T-3'
BAC36-1R	5'-CGC CGT AGT ACG CTT TGT T-3'

Sonde 2:

BAC36-2F	5'-GGC ACG AAC TTC TTC CGT A-3'
BAC36-2R	5'-GGT TCC GGA TCA ATT CTC AT-3'

Sonde 3:

BAC36-3F	5'-CGT CGA CAG CGA CAC ACT T-3'
BAC36-3R	5'-CGT CAT CTG CAT CAA GAA CTA GT-3'

Sonde 4:

BAC36-4F	5'-GCT TCA AAT ATC ACC CCA GCT-3'
BAC36-4R	5'-GTG CCT GGA GAA ATG CTC TT-3'

Sonde 5:

BAC36-5F	5'-CGT TGA TCA ATG GCC TGT T-3'
BAC36-5R	5'-GCT GGT CAT GAA ATG ACG AA-3'

Sonde 6:

BAC36-6F	5'-CCA GCG TTT ATG CAC TGG TT-3'
BAC36-6R	5'-CAA CCC ATA TCT CAT TCC CTT-3'

Sonde 7:

BAC36-7F	5'-GAC TGT GCG CAC GTT TGA T-3'
BAC36-7R	5'-GCT GGC TGT TAT CCA CTG AGA A-3'

Sonde 8:

BAC36-8F	5'-GCA TAT CCA CTC AGT TCC ACA TT-3'
BAC36-8R	5'-GGC CGT AAT ATC CAG CTG AA-3'

Sonde 9:

BAC36-9F	5'-CAG CTG GAT ATT ACG GCC TT-3'
BAC36-9R	5'-CGG CGT TTA AAT AAT CTC GTT A-3'

4.3.5 Verbrauchsmaterialien

Sterilfiltration:

Schleicher & Schüll, Membranfilter RC58

Faltenfilter:

Macherey & Nagel, Nucleobond folded filter

Chromatographiepapier:

Whatman 3MM

Elektroporation:

BIO-RAD, Elektroporationsküvetten 0,2 cm

Gelextraktion:

QIAGEN, QIAex II Gel Extraction Kit und QIAquick Gel Extraction Kit

DNA-Isolierung:

Macherey & Nagel, Nucleobond AX100 und Nucleobond AX500

QIAGEN, Tip 500 und QIAprep spin Miniprep Kit

MILLIPORE, Multiscreen Kit

4.4 Allgemeine Grundtechniken

4.4.1 Sterilisieren

Die Sterilisation von Nährmedien und Pufferlösungen erfolgte durch Autoklavieren bei 121 °C für 20 min oder durch Filtration (0,22 µm). Glasgeräte wurden bei 180 °C für 3 h sterilisiert.

4.4.2 Herstellung von Agarplatten

Dem verwendeten Nährmedium werden 15 g Agar beigemischt. Nach dem Autoklavieren ließ man das Medium auf eine Temperatur von 50 – 60 °C abkühlen, bevor das entsprechende Antibiotikum in einem Volumenverhältnis von 1/1000 zugegeben wurde. Für die blau / weiß-Selektion wurden zusätzlich 70 mg X-Gal (in 300 µl DMF gelöst) und 80 µl 1 M IPTG pro Liter Medium zugegeben. Für die Herstellung von Hybridisierungsfiltren waren luftblasenfreie Agarplatten notwendig. Eventuell vorhandene Luftblasen wurden gleich nach dem Gießen der Platten durch Abflammen mit einem Bunsenbrenner entfernt.

4.4.3 Herstellung von kompetenten Zellen

Mit einer Einzelkolonie wurden 50 ml LB-Medium angeimpft. Diese Vorkultur wurde über Nacht bei 37°C unter schütteln inkubiert. Anschließend wurden 1 Liter, auf 37°C vorgewärmtes LB-Medium mit 25 ml Vorkultur beimpft. Die Zellen wurden bei 37°C unter Schütteln bis zu einer OD600 von 0,5-1 wachsen gelassen. Nachdem die Kultur auf 4 Zentrifugenbecher verteilt war wurde diese für 15 min auf Eis gekühlt bevor die Zellen durch Zentrifugation für 12 min in einem vorgekühlten GS3-Rotor bei 5000 Upm abgetrennt wurden. Die Becher wurden mit einigen Millilitern sterilem, 4°C kaltem 10%igen Glycerin vorsichtig gewaschen ohne das Zellpellet zu resuspendieren. Die Resuspendierung der Zellpellets erfolgt mit jeweils 200ml kaltem 10%igen Glycerin und anschließend wurde die Zentrifugation wiederholt. Nach diesem Zentrifugationsschritt wurden die Pellets in je 100 ml kaltem 10%igen Glycerin aufgenommen, der Inhalt von 2 Bechern vereinigt und wiederum wie oben zentrifugiert. Nun wurden die Pellets in 20 ml kaltem 10%igen Glycerin resuspendiert, in ein 50 ml Zentrifugenröhrchen überführt und für 12 min in einem vorgekühlten SS34-Rotor bei 6000 Upm zentrifugiert. Danach wurde das Pellet in einem Endvolumen von 1 ml kaltem 10%igem Glycerin resuspendiert. Die so aufgenommenen Zellen werden in 50 µl Aliquots in flüssigem Stickstoff schockgefroren und bei -70°C gelagert.

4.4.4 Elektrotransformation von *E. coli*

Mit der Ligationslösung (5-15 µl) wurde zur Steigerung der Transfektionseffizienz eine thermische Ligasedenaturierung durchgeführt (Ymer, 1991). Dazu wurde der Ligationsansatz für 10 min bei 65 °C inkubiert und anschließend auf Eis gekühlt.

Ein 50 µl Aliquot der elektrokompenten Zellen wurde nach dem Auftauen auf Eis mit der DNA-Lösung vermischt und in eine eiskalte, sterile 0,2 cm Elektroporationsküvette überführt. Die Elektroporation erfolgte bei 2500 Volt, 25 µFarad, 200 Ohm und einer Pulsdauer von etwa 4,5 ms (Fromm *et al.*, 1987). Nach sofortiger Zugabe von 1 ml bzw. 4 ml (Herstellung von "shotgun"-Banken) SOC-Medium wurden die Kulturen 45 min bei 37 °C inkubiert. Anschließend wurde die Kultur mit einem Drigalsky-Spatel (20-200µl) auf LB-Agar-Platten (100 µg/ml Ampicillin) ausgestrichen.

4.4.5 Lagerung von Bakterienkolonien

Für die kurzfristige Lagerung von Bakterienkolonien wurden diese auf Agarplatten mit entsprechendem Selektionsmedium ausplattiert und über Nacht bei 37°C inkubiert. Die Lagerung erfolgte bei 4°C.

Für die langfristige Lagerung von Bakterienkolonien wurde ein Aliquot einer frischen Übernachtskultur im Verhältnis 1:1 mit 87%igem, sterilen Glycerin vermischt. Die Lagerung dieser Glycerinkulturen erfolgte bei –70°C.

Die langfristige Lagerung von "shotgun"-Banken wurden die Bakterienkolonien in einem speziellen Nährmedium angezogen, das sich zum Einfrieren der Zellen bei –70°C eignete. Zu 2YT-Medium wurden 10% Gefriermedium und Antibiotika zugegeben. Mit diesem Medium wurden die Mikrotiterplatten befüllt und angeimpft. Nach einer Inkubation im Brutschrank bei 37°C über Nacht konnten die Platten eingefroren werden.

4.4.6 DNA-Fällung

Falls nicht gesondert erwähnt wurde für die Fällung von DNA Isopropanol bzw. PEG-Lösung eingesetzt.

Für größere Volumina (ab 5 ml) wurden zu der wässrigen DNA-Lösung 0,7 Volumen Isopropanol zugegeben und für 30 min bei RT inkubiert. Für Volumina bis 25 ml wurden für das Abzentrifugieren des Präzipitats 50 ml Zentrifugenröhrchen verwendet. Die Zentrifugation erfolgte in einem vorgekühlten SS34-Rotor bei 4°C und 15000 Upm für 30min. Nach vorsichtigem Abgießen des Überstands wurde mit 10 ml gekühltem 80 %igem Ethanol gewaschen.

Anschließend wurde unter gleichen Bedingungen für 20 min zentrifugiert. Nach dem Abgießen des Überstands wurde die DNA für 5 min bei RT getrocknet und dann in 200 µl H₂O aufgenommen.

Für kleinere Volumina (bis 500µl) wurde für die Fällung der DNA PEG-Lösung eingesetzt. Die Zugabe der PEG-Lösung erfolgte im Verhältnis 1:1. Nach gründlichem Vermischen folgte eine Inkubation für 10 min bei RT. Zur Abtrennung des Präzipitats wurde für 10 min bei RT und 14000 Upm zentrifugiert und anschließend der Überstand vorsichtig abpipettiert. Das Waschen der DNA erfolgte mit 1 ml gekühltem 80%igem Ethanol und anschließendem Zentrifugieren unter gleichen Bedingungen wie oben beschrieben. Nach dem Trocknen der DNA für ca. 5 min wurde diese in H₂O (6 – 30 µl) aufgenommen.

4.4.6.1 Ermittlung der Fällungsbedingungen zur PCR-Produktaufreinigung

Standardmäßig wurden 5 pmol je Primer in eine PCR mit einem Reaktionsvolumen von 20 µl eingesetzt. Das entspricht einer Gesamtprimerkonzentration von 0,5 pmol/µl. Aus einer PCR mit Bakterienklonen in einem Reaktionsvolumen von 20 µl konnten durchschnittlich 750 ng eines etwa 1,3 kbp großen PCR-Produkts gewonnen werden. Das entspricht einer Konzentration von 37,5 ng/µl DNA. Zur Ermittlung der Fällungsbedingungen zur PCR-Produktaufreinigung wurden jeweils von den Primern bzw. der DNA mit diesen Konzentrationen Stammlösungen mit 1x Taq-Puffer hergestellt. Jeweils 50 µl dieser Lösungen wurden in die Fällung eingesetzt. Diese erfolgte durch Zugabe von 0,75 Volumen Isopropanol bzw. durch Zugabe von 3 Volumen Ethanol (75% Endkonzentration). Im Anschluss folgte eine einheitliche Inkubation für 20 min bei Raumtemperatur. Dann wurden die Proben für 20, 30 bzw. 40 min bei 2550 x g zentrifugiert. Nach abpipettieren des Überstands wurde mit 75 µl 80%igem Ethanol gewaschen und nochmals für 10 min bei 2550 x g zentrifugiert. Der Überstand wurde wieder abpipettiert und anschließend wurden die Proben bei Raumtemperatur getrocknet. Das Lyophilisat wurde in 80 µl H₂O aufgenommen und die Konzentration UV-spektrometrisch (Kapitel 4.4.8) bestimmt.

4.4.7 Aufreinigung von DNA mit Cäsiumchlorid-Gradienten

Zur Aufreinigung von DNA mittels Cäsiumchlorid-Gradienten wurden 10,788 ml DNA-Isolat mit 12,354 g Cäsiumchlorid und 1,5 ml Ethidiumbromid (10 mg/ml) vermischt. Zum Lösen des Cäsiumchlorids wurde für mehrere Stunden bei 8°C inkubiert. Zum Entfernen der unlösten Rückstände wurde für 20 min bei 6000 Upm zentrifugiert und der Überstand in ein Ultrazentrifugationsröhrchen eingeschweißt. Die Zentrifugation erfolgte bei 44000 Upm und

18°C für 48 h in einem Ti 70-Rotor. Anschließend wurden unter UV-Licht ca. 1-2 ml Ethidiumbromid-DNA-Lösung abgenommen und in ein 10 ml Kulturröhrchen überführt. Zur Entfernung des Ethidiumbromids wurde mit 1 Volumen Isoamylalkohol (H₂O-gesättigt) ausgeschüttelt und der Überstand abpipettiert. Dieser Vorgang wurde so oft wiederholt, bis keine Rotfärbung mehr erkennbar war. Dann wurde die so aufgereinigte DNA-Lösung in ein neues Kulturröhrchen überführt, mit 500 µl 3 M Natriumacetat und 3,5 ml Ethanol versetzt und über Nacht bei 4°C inkubiert. Die Zentrifugation erfolgte 4000 Upm für 10 min. Nach Abgießen des Überstands wurde 3 mal mit je 500 µl 70%igem Ethanol gewaschen. Die Zentrifugation erfolgte wie oben. Der Überstand wurde verworfen und das DNA-Pellet bei RT getrocknet. Danach wurde die DNA in 250 µl H₂O aufgenommen, in ein 1,5 ml Eppendorf-Gefäß pipettiert, mit 125 µl Amoniumacetat und 600 µl Isopropanol versetzt und für 10 min bei RT inkubiert. Nach der Zentrifugation für 10 min bei 12000 Upm folgte ein Waschschriff mit 100 µl 70%igem Ethanol und Wiederholung der Zentrifugation wie zuvor. Das getrocknete DNA-Pellet wurde in 50 µl TE-Puffer aufgenommen.

4.4.8 UV-spektrometrische Konzentrationsbestimmung von DNA

Die Bestimmung der OD_{260 nm} erfolgte an einem Spektrophotometer mit variabler Wellenlänge in 1 cm Quarzglasküvetten. Ein Aliquot der DNA-Lösung (5-20 µL) wurde zu einem Gesamtvolumen von 100 µl mit H₂O aufgefüllt und gegen H₂O vermessen. Das Pharmacia-Gerät (Gene Quant) berechnete die DNA-Konzentrationen automatisch.

4.5 Isolierung von DNA

4.5.1 Analytische Isolierung von Plasmiden

Puffer P1 :	25 mM Tris-HCl pH 8,0; 10 mM EDTA; 100 µg/ml RNaseA; 50 mM Glucose.
Puffer P2 :	0,2 M NaOH; 1 % (w/v) SDS.
Puffer N3 :	3 M Kaliumacetat; pH 4,8.
Puffer PE :	50 mM Tris-HCl; pH 8,0; 0,1 M Natriumacetat, EtOH (siehe Herstellerangabe)

Plasmid-DNA in analytischen Mengen (bis etwa 10 µg) wurde aus stationären LB-Übernachtskulturen mit 4 ml Volumen isoliert. Das Zellpellet wurde in 250 µl P1-Puffer (100 µg/ml

RNase A) aufgenommen und mit 250 µl P2-Puffer durch alkalische SDS-Lyse (5 min/RT) aufgeschlossen (Birnboim & Doly, 1979). Durch die Neutralisation mit 350 µl N3-Puffer wurden die Zelltrümmer, die Proteine und teilweise auch die genomische DNA präzipitiert. Sie wurden durch eine Zentrifugation bei 14000 U/min (15 min/RT) sedimentiert.

Zur vollständigen Aufreinigung wurde der klare Überstand auf eine Anionenaustauschersäule (QIAprep Spin Plasmid-Kit der Firma Qiagen) geladen und bei 9700 Upm (60 s/ RT) zentrifugiert. Um eine Erhöhung der Ausbeute zu erzielen wurde dieser Schritt wiederholt. Es folgte ein Waschschriff mit 750 µl PE-Puffer . Nach der Zentrifugation wurde das Auffanggefäß entleert und erneut zentrifugiert um eine möglichst vollständige Entfernung des PE-Puffers zu erzielen. Anschließend wurde die Plasmid-DNA mit zweimal 30 - 50 µl auf 65°C vorgewärmten H₂O durch Zentrifugation von der Säule eluiert.

4.5.2 Isolierung von Plasmiden im 96er-Raster für Sequenzierung

Lösung 1:	30 mM Glucose; 15 mM Tris-HCl, pH 8,0; 30 mM EDTA; 60 µg/ml RNase A
Lösung 2:	0,2 M NaOH; 1% (w / v) SDS
Lösung 3:	3,6 M Kaliumacetat; 2,4 M konzentrierte Essigsäure
Bindelösung:	6,1 M KI

Ein Block mit 96, je 2 ml fassenden Kammern (Qiagen „Square-Well-Block“) wurde pro Kammer mit 1,2 ml Antibiotikaversetztem 2YT-Medium befüllt. Dann wurden mittels eines 96er-Stempels Bakterienklone aus einer 384er-Mikrotiter-Dauerkulturplatte überimpft und der Block mit einer luftdurchlässigen Folie verschlossen. Anschließend ließ man die Bakterienklone bei 37°C unter starkem Schütteln (230 U/min) für 18-20 h über Nacht wachsen. Das Abzentrifugieren der Zellen erfolgte bei 1500 x g für 5 min in einer Mikrotiterplatten-Zentrifuge (Juon). Die vollständige Resuspendierung der Zellen mit je 80 µl Lösung 1 erfolgte durch Vortexen. Nach Zugabe von je 80 µl Lösung 2 wurde auf einem Vortexer mit Mikrotiterplatten-Halterung bei maximaler Geschwindigkeit für 1 min bei RT durchmischt und anschließend für weitere 4 min inkubiert. Es wurden je 80 µl Lösung 3 zugegeben und wie zuvor auf dem Vortexer für 2 min durchmischt. Die FB-Platte, an der die DNA gebunden und gereinigt wurde, wurde in der Vakuumkammer positioniert und die NA-Platte zur Filterung der Lysate wurde in die dafür vorgesehene Öffnung im Deckel der Vakuumkammer darüber gestellt. Nach Überführung der Lysate in die NA-Platte wurde durch Anlegen eines Vakuums von 10 in Hg für 3 min die Lösung in die FB-Platte gefiltert. Anschließend wurde die FB-

Platte in den Deckel der Vakuumkammer gestellt und je 150 µl Bindelösung zugegeben. Mit maximalem Vakuum für 1 min wurde die Lösung durchgesaugt und die DNA auf dem Säulenmaterial gebunden. Es folgten 2 Waschschrte mit jeweils 200 µl 80%igem Ethanol bei maximalem Vakuum. Beim ersten Waschschrte für 1 min und beim zweiten für 3 min. Danach wurde die FB-Platte für 1 h bei 65°C getrocknet um Ethanolreste zu entfernen. Danach wurde in die Vakuumkammer eine 96er Mikrotiterplatte und auf die Öffnung in deren Deckel die FB-Platte gestellt. Die Eluation der DNA in die Mikrotiterplatte erfolgte 2 x mit je 35 µl auf 37°C temperiertem H₂O bei maximalem Vakuum für 1 bzw. 3 min.

4.5.3 Präparative Isolierung von Plasmiden

Puffer P1 :	50 mM Tris-HCl; pH 8,0; 10 mM EDTA; 100 µg/ml RNaseA.
Puffer P2 :	0,2 M NaOH; 1 % SDS.
Puffer P3 :	3,0 M Kaliumacetat; pH 5,5.
Puffer QBT :	750 mM NaCl; 50 mM MOPS, 15 % EtOH; 0,15 % Triton X-100; pH 7,0.
Puffer QC :	1,0 M NaCl; 50 mM MOPS; 15 % EtOH; pH 7,0.
Puffer QF :	1,25 mM NaCl; 50 mM Tris-HCl; pH 8,5; 15 % EtOH.

Die Isolierung der Plasmid-DNA erfolgte unter Verwendung des Qiagen Maxi-Kits. Die zu präparierenden Bakterienklone wurden in 500 ml LB-Medium in einer Schüttelkultur bei 37 °C über Nacht inkubiert. Nach dem Abzentrifugieren (15 Minuten, 5000 Upm) wurde das Pellet mit 10 ml P1-Puffer bei RT im Schüttler resuspendiert. Nach Zugabe von 10 ml P2-Puffer wurde der Zentrifugenbecher zur Durchmischung vorsichtig 6–8 mal gewendet und anschließend höchstens 5 min inkubiert. Nach Zugabe von 10 ml P3-Puffer (RT) und vorsichtigem 6–8maligem Umschwenken erfolgte eine 30 min Inkubation auf Eis. Abschließend wurde 30 min bei 4°C/15000 Upm in einem vorgekühlten SS34-Rotor zentrifugiert. Die Säulen wurden mit 10 ml QBT - Puffer equilibriert, die Faltenfilter mit H₂O befeuchtet und der klare Überstand nach der Zentrifugation über den Filter gegeben auf eine Anionenaustauscher-Säule gegeben. . Danach wurde zweimal mit 30 ml QC-Puffer gewaschen und abschließend mit 15 ml, auf 50°C vorgewärmten QF-Puffer eluiert. Die Entsalzung erfolgte durch Isopropanol-Fällung (Kapitel 5.4.6). Der getrocknete Niederschlag wurde in 200-1000 µl H₂O aufgenommen. Die Konzentrationsbestimmung erfolgte am Photometer.

4.5.4 Präparative Isolierung von Cosmid-, BAC- und PAC-DNA

4.5.4.1 Fällungsverfahren

10 x GET-Puffer: 50 mM Glucose, 25 mM Tris-HCl, 10 mM EDTA, pH 8,0

Lösung 1: 0,2 M NaOH, 1% SDS (w / v)

Lösung3: 3 M Kaliumacetat, pH 5,1

5 ml 2YT-Medium mit entsprechendem Antibiotikum wurden morgens mit einer Einzelkolonie angeimpft und für ca. 8 h unter starkem Schütteln bei 37°C inkubiert. Von dieser Vorkultur wurden 3 Liter 2YT-Medium mit Antibiotikum mit 1 ml je Liter angeimpft. Die Inkubation erfolgte bei 37°C und 200 Upm über Nacht. Die Übernachtskultur wurde zu je 250 ml auf eine entsprechende Anzahl Zentrifugenbecher verteilt und bei 4°C und 4000 Upm in einem GSA-Rotor für 10 min zentrifugiert. Nach abgießen des Überstands wurden die Zellpellets in je 6 ml 1 x GET-Puffer und 30 mg Lysozym vorsichtig resuspendiert und vereinigt. Von der Zellsuspension wurden je 12 ml in eine entsprechende Anzahl 50 ml Zentrifugenröhrchen verteilt und für 5 min auf Eis gestellt. Es folgte die Zugabe von je 14 ml Lösung 1. Nach vorsichtigen Mischen durch umschenken wurde für weitere 5 min auf Eis inkubiert und anschließend je 12 ml Lösung 2 zugegeben. Die Proben wurden für 30 min auf Eis gestellt, dann bei 4°C und 10000 Upm in einem SS34-Rotor für 15 min zentrifugiert und dann der klare Überstand (ca. 25 ml) in 50 ml Corning-Röhrchen überführt. Nach Zugabe von 25 ml Isopropanol wurde für 2 h bei RT inkubiert und anschließend bei 4000 Upm für 25 min im Swingout-Rotor (Haereus) zentrifugiert. Es folgte ein Waschschriff mit 15 ml 70%igem Ethanol. Es wurde für 15 min bei ansonsten gleichen Bedingungen wie zuvor zentrifugiert, anschließend der Überstand abgegossen und danach die Pellets für 30 min getrocknet. Diese wurden dann in je 2 ml TE-Puffer und 25 µl RNase A (50 mg / ml) gelöst und bei 4°C über Nacht inkubiert. Nach dem Vereinigen aller Isolate in einem Corning-Röhrchen wurde 1 Volumen Phenol zugegeben, vorsichtig durchmischt und dann für 15 min bei 4000 Upm im Swingout-Rotor (Haereus) zentrifugiert. Der Überstand wurde vorsichtig abpipettiert und in ein neues Corning-Röhrchen überführt. Dieser Vorgang wurde mit jeweils 1 Volumen Phenol/Chloroform und Chloroform wiederholt. Für die Fällung wurden 2 Volumen Ethanol zugegeben, für 2 h bei -20°C inkubiert und für 20 min bei 4000 Upm im Swingout-Rotor (Haereus) zentrifugiert. Anschließend wurde der Überstand abgegossen, das Präzipitat für 30 min bei RT getrocknet und dann in 11 ml TE-Puffer aufgenommen.

Die weitere Aufreinigung erfolgte durch CsCl-Gradienten (Kapitel 5.4.7) und anschließender Behandlung mit „Plasmid Safe“ (Kapitel 5.6.2). Nach der photometrischen Bestimmung der Konzentration konnte die DNA zur Herstellung von "shotgun"-Banken eingesetzt werden.

4.5.4.2 Säulenchromatographieverfahren

4.5.4.2.1 Qiagen Maxi-Kit

Puffer P1 :	50 mM Tris-HCl; pH 8,0; 10 mM EDTA; 100 µg/ml RNaseA.
Puffer P2 :	0,2 M NaOH; 1 % SDS.
Puffer P3 :	3,0 M Kaliumacetat; pH 5,5.
Puffer QBT :	750 mM NaCl; 50 mM MOPS, 15 % EtOH; 0,15 % Triton X-100; pH 7,0.
Puffer QC :	1,0 M NaCl; 50 mM MOPS; 15 % EtOH; pH 7,0.
Puffer QF :	1,25 mM NaCl; 50 mM Tris-HCl; pH 8,5; 15 % EtOH.

Dieses beschriebenen Verfahren entspricht einem modifiziertem Standardprotokoll. Um ausreichend DNA zu erhalten wurden 3 Liter 2YT-Medium mit entsprechendem Antibiotikum mit einer Glycerinkultur (25µl / L) beimpft und bei 37°C über Nacht geschüttelt. Anschließend wurden die Zellen durch mehrmaliges Zentrifugieren bei 4°C und 5700 Upm für 12 min abgetrennt. Das Pellet wurde in 42 ml P1-Puffer bei RT im Schüttler resuspendiert. Nach Zugabe von 42 ml P2-Puffer wurde der Zentrifugenbecher zur Durchmischung vorsichtig 6–8 mal gewendet und anschließend höchstens 5 min inkubiert. Danach wurden 42 ml P3-Puffer (RT) zugegeben und durch vorsichtige 6–8maligem Umschwenken durchgemischt. Es folgte eine Inkubation für 10 min auf Eis. Abschließend wurde 50 min bei 4°C / 9000 Upm in einem vorgekühlten GS3-Rotor zentrifugiert. Der Überstand wurde zur Entfernung von Schwebstoffen durch befeuchtete Faltenfilter in kleinere Zentrifugen-Becher filtriert. Im Anschluss folgte eine Fällung der DNA durch Zugabe von 90 ml Isopropanol und Zentrifugation bei 4°C / 13000 Upm für 50 min in einem vorgekühltem GSA-Rotor. Nach Abgießen des Überstands wurde für 5 min bei RT getrocknet. Das Präzipitat wurde in 2 ml TE-Puffer aufgenommen und der Becher mit 10 ml QBT-Puffer nachgespült. Beide Chargen wurden vereinigt. Vor der Beschickung der Anionenaustauscher-Säule (Tip 500) wurde diese mit 5 ml QBT - Puffer equilibriert. Danach wurde mit je 12 ml QC-Puffer dreimal gewaschen und abschließend zweimal mit je 6 ml QF-Puffer eluiert. Die Entsalzung erfolgte durch Isopropanol-Fällung

(Kapitel 5.4.6). Der getrocknete Niederschlag wurde in 800 µl TE-Puffer aufgenommen. Die Konzentrationsbestimmung erfolgte am Photometer.

4.5.4.2.2 Macherey & Nagel-Kit Variante 1

S1-Puffer: 50 mM Tris / HCl, 10 mM EDTA, 100 µg RNase A /ml, pH 8,0

S2-Puffer: 200 mM NaOH, 1% SDS

S3-Puffer: 2,80 M KAc, pH 5,1

N2-Puffer: 100 mM Tris, 15% Ethanol, 900 mM KCl, mit H₃PO₄ auf pH 6,3 eingestellt

N3-Puffer: 100 mM Tris, 15% Ethanol, 1150 mM KCl, mit H₃PO₄ auf pH 6,3 eingestellt

N5-Puffer: 100 mM Tris, 15% Ethanol, 1000 mM KCl, mit H₃PO₄ auf pH 6,5 eingestellt

Dieses beschriebenen Verfahren entspricht einem modifiziertem Standardprotokoll. Um ausreichend DNA zu erhalten wurden 3 L 2YT-Medium mit entsprechendem Antibiotikum mit einer Glycerinkultur (25µl / L) beimpft und bei 37°C über Nacht geschüttelt. Anschließend wurden die Zellen durch mehrmaliges Zentrifugieren bei 4°C / 5700 Upm für 12 min abgetrennt. Das Pellet wurde in 42 ml S1-Puffer bei RT im Schüttler resuspendiert. Nach Zugabe von 42 ml S2-Puffer wurde der Zentrifugenbecher zur Durchmischung vorsichtig 6–8 mal gewendet und anschließend höchstens 5 min inkubiert. Danach wurden 42 ml S3-Puffer (RT) zugegeben und durch vorsichtige 6–8maligem Umschwenken durchmischt. Es folgte eine Inkubation für 10 min auf Eis. Abschließend wurde 50 min bei 4°C / 9000 Upm in einem vorgekühlten GS3-Rotor zentrifugiert. Der Überstand wurde zur Entfernung von Schwebstoffen durch befeuchtete Faltenfilter in kleinere Zentrifugen-Becher filtriert. Im Anschluss folgte eine Fällung der DNA durch Zugabe von 90 ml Isopropanol und Zentrifugation bei 4°C / 13000 Upm für 50 min in einem vorgekühltem GSA-Rotor. Nach Abgießen des Überstands wurde für 5 min bei RT getrocknet. Das Präzipitat wurde in 2 ml TE-Puffer aufgenommen und der Becher mit 10 ml N2-Puffer nachgespült. Beide Chargen wurden vereinigt. Vor der Beschickung der Anionenaustauscher-Säule (AX 500) wurde diese mit 5 ml N2 - Puffer equilibriert. Danach wurde mit je 12 ml N3-Puffer dreimal gewaschen und abschließend zweimal mit je 6 ml N5-Puffer eluiert. Die Entsalzung erfolgte durch Isopropanol-Fällung (Kapitel 5.4.6). Der getrocknete Niederschlag wurde in 800 µl TE-Puffer aufgenommen. Die Konzentrationsbestimmung erfolgte am Photometer.

4.5.4.2.3 Macherey & Nagel-Kit Variante 2

S1-Puffer:	50 mM Tris / HCl, 10 mM EDTA, 100 µg RNase A /ml, pH 8,0
S2-Puffer:	200 mM NaOH, 1% SDS
S3-Puffer:	2,80 M KAc, pH 5,1
N2-Puffer:	100 mM Tris, 15% Ethanol, 900 mM KCl, mit H ₃ PO ₄ auf pH 6,3 eingestellt
N3-Puffer:	100 mM Tris, 15% Ethanol, 1150 mM KCl, mit H ₃ PO ₄ auf pH 6,3 eingestellt
N5-Puffer:	100 mM Tris, 15% Ethanol, 1000 mM KCl, mit H ₃ PO ₄ auf pH 6,5 eingestellt

500 ml 2YT-Medium mit entsprechendem Antibiotikum wurde mit 25 µl einer Glycerinkultur beimpft und bei 37°C über Nacht geschüttelt. Nach der Zentrifugation bei 4°C und 6000 Upm für 15 min wurde das Pellet in 24 ml S1-Puffer bei RT im Schüttler resuspendiert. Nach Zugabe von 24 ml S2-Puffer wurde der Zentrifugenbecher zur Durchmischung vorsichtig 6–8 mal gewendet und anschließend höchstens 5 min inkubiert. Nach Zugabe von 24 ml S3-Puffer (RT) und vorsichtigem 6–8maligem Umschwenken erfolgte eine 50 min Inkubation auf Eis. Abschließend wurde 50 min bei 4°C/13000 Upm in einem vorgekühlten GSA-Rotor zentrifugiert. Die Säulen (AX 500) wurden mit 5 ml N2 - Puffer equilibriert, die Faltenfilter mit H₂O befeuchtet und der klare Überstand nach der Zentrifugation über den Filter gegeben auf eine Anionenaustauscher-Säule gegeben. Der Durchlauf wurde anschließend erneut auf die Säule gegeben. Danach wurde mit je 12 ml N3-Puffer dreimal gewaschen und abschließend zweimal mit je 6 ml, auf 50°C vorgewärmten N5-Puffer eluiert. Die Entsalzung erfolgte durch Isopropanol-Fällung (Kapitel 5.4.6). Der getrocknete Niederschlag wurde in 200 µl H₂O aufgenommen. Die Konzentrationsbestimmung erfolgte am Photometer.

4.5.4.2.4 Macherey & Nagel-Kit Variante 3

S1-Puffer:	50 mM Tris / HCl, 10 mM EDTA, 100 µg RNase A /ml, pH 8,0
S2-Puffer:	200 mM NaOH, 1% SDS
S3-Puffer:	2,80 M KAc, pH 5,1
N2-Puffer:	100 mM Tris, 15% Ethanol, 900 mM KCl, mit H ₃ PO ₄ auf pH 6,3 eingestellt
N3-Puffer:	100 mM Tris, 15% Ethanol, 1150 mM KCl, mit H ₃ PO ₄ auf pH 6,3 eingestellt
N5-Puffer:	100 mM Tris, 15% Ethanol, 1000 mM KCl, mit H ₃ PO ₄ auf pH 6,5 eingestellt

150 ml 2YT-Medium mit entsprechendem Antibiotikum wurde mit 25 µl einer Glycerinkultur beimpft und bei 37°C über Nacht geschüttelt. Nach der Zentrifugation bei 4°C und 6000 Upm

für 15 min wurde das Pellet in 30 ml S1-Puffer bei RT im Schüttler resuspendiert. Nach Zugabe von 30 ml S2-Puffer wurde der Zentrifugenbecher zur Durchmischung vorsichtig 6–8 mal gewendet und anschließend höchstens 5 min inkubiert. Nach Zugabe von 30 ml S3-Puffer (RT) und vorsichtigem 6–8maligem Umschwenken erfolgte eine 50 min Inkubation auf Eis. Abschließend wurde 50 min bei 4°C/13000 Upm in einem vorgekühlten GSA-Rotor zentrifugiert. Die Säulen (AX 100) wurden mit 2 ml N2 - Puffer equilibriert, die Faltenfilter mit H₂O befeuchtet und der klare Überstand nach der Zentrifugation über den Filter gegeben auf eine Anionenaustauscher-Säule gegeben. Der Durchlauf wurde anschließend erneut auf die Säule gegeben. Danach wurde mit je 4 ml N3-Puffer dreimal gewaschen und abschließend zweimal mit je 3 ml, auf 50°C vorgewärmten N5-Puffer eluiert. Die Entsalzung erfolgte durch Isopropanol-Fällung (Kapitel 5.4.6). Der getrocknete Niederschlag wurde in 200 µl H₂O aufgenommen. Die Konzentrationsbestimmung erfolgte am Photometer.

4.5.5 Isolierung von DNA aus Agarosegelen

Zur Isolierung und Reinigung von DNA-Fragmenten aus Agarosegelen wurde der QIAquick-Gel Extraction-Kit von Qiagen verwendet. Mit der Methode von Qiagen konnte das Gel durch Inkubation bei 50°C für 5 min mit NaI-Lösung (QX1-Puffer) aufgelöst und die DNA nach Auftrag auf die Anionentauschersäule chromatographisch aufgereinigt werden. Es wurde mit 500 µL QX1-Puffer und mit 750 µL PE-Puffer gewaschen. Die Eluation der DNA erfolgte mit 50 µL H₂O.

Zur Isolierung und Reinigung von DNA-Fragmenten aus Agarosegelen, die kleiner als 500 bp waren, wurde der QIAEX-Kit von Qiagen verwendet, um eine höhere Ausbeute (ca.70-80 %) zu erhalten. Mit der Methode von Qiagen konnte das Gel durch Inkubation bei 50°C für 5 min mit NaI-Lösung (QX1-Puffer) aufgelöst und die DNA in 5 min bei RT an einen Anionentauscher (QIAEX-Glasmilch) gebunden werden. Es folgte ein jeweils zweimaliges Waschen mit ethanolischen, salzhaltigen Puffern (QX2, QX3) und nach kurzem Trocknen bei RT die Eluation der DNA mit 2 x 10-20 µL H₂O (je 5 min, RT). Abweichend von der Herstellervorschrift wurde im allgemeinen das Doppelte der angegebenen Menge des Ionentauschers eingesetzt.

4.6 Modifizierung von DNA

4.6.1 DNA-Verdau mit Restriktionsenzymen

Der analytische Restriktionsverdau wurde mit 200-800 ng DNA und 2-3 Einheiten (U) der jeweiligen Restriktionsenzyme in maximal 25 µl durchgeführt. Die Inkubationszeit bei 37 °C variierte von mindestens 1-2 h.

Der präparative Restriktionsverdau wurde mit 5-25 µg DNA und 5 Einheiten (U) Enzym pro µg DNA in maximal 300 µl durchgeführt. Die Inkubationszeit bei 37 °C betrug in diesem Fall bis zu 3h. Dabei wurden immer die vom Hersteller für die Restriktionsenzyme empfohlenen Puffer verwendet. War dies bei gewissen Enzymkombinationen nicht möglich, wurde der für die jeweilige Enzymkombination sinnvollste Puffer verwendet. Im Anschluss wurde mit der Agarose-Gelelektrophorese eine Restriktionsanalyse oder eine Isolierung und Aufreinigung der DNA-Fragmente vorgenommen.

4.6.2 Behandlung von DNA mit ATP-abhängiger DNase

10x Reaktionspuffer: 330 mM Tris-acetat (pH 7,8), 660 mM Kaliumacetat,
 100 mM Magnesiumacetat, 5 mM DTT.

Für diesen Schritt wurde die ATP-abhängige DNase „Plasmid Safe“ der Fa. Biozym verwendet. Zu 205 µl DNA-Lösung wurden 25 µl 10x Reaktionspuffer, 10 µl 25 mM ATP und 10 µl „Plasmid Safe“ (10 U/µl) gegeben. Nach vollständiger Durchmischung folgte eine Inkubation bei 37 °C für 2 Stunden. Anschließend wurde das Enzym durch Inkubation bei 70 °C für 30 min denaturiert. Die nachfolgende Aufreinigung der DNA erfolgte durch PEG-Fällung (Kapitel 5.4.6).

4.6.3 Fragmentierung von Cosmid-, BAC- und PAC-DNA

4.6.3.1 Ultraschallfragmentierung

Für die Fragmentierung der DNA durch Ultraschall wurden 30-50 µg in einem Gesamtvolumen von 100 µl TE-Puffer gelöst. Die DNA wurde insgesamt mit 11 Pulsen mit einer Intervalldauer von 0,2 s bei „Powerlevel“ 30 der Stufe „low“ beschallt. Das 1,5 ml Eppendorf-Gefäß wurde dabei so gehalten, dass die Stabsonde möglichst tief in die Lösung ragte ohne die Wandungen des Gefäßes zu berühren. Nach jeweils 3 Pulsen wurde auf Eis gekühlt. An-

schließlich wurden 20 µl Gelloading-Puffer zugegeben und die gesamte Menge wurde auf 4 Spuren eines Agarosegels aufgeteilt. Die Auftrennung der Fragmente erfolgte bei 60 V für 3-4 h. Die Fragmente der Größen 1-1,3 kbp und 2-2,8 kbp wurden aus dem Gel ausgeschnitten und extrahiert (Kapitel 5.5.3).

4.6.3.2 Fragmentierung durch hydrodynamisches Scheren

Für die Herstellung von DNA-Fragmenten der Größen 1,3-1,5 kbp und 2,5-3 kbp müssen bei diesem Verfahren 2 Schergänge durchgeführt werden. Aus diesem Grund wurden 8-15 µg DNA in 200 µl H₂O gelöst. Die Entfernung von Schwebstoffen erfolgte durch Zentrifugieren für 10 min bei 14000 Upm (RT) und anschließendem Überführen des Überstands in ein neues 1,5 ml Eppendorf-Gefäß. Von der so vorbereiteten Proben wurden pro Schergang 100 µl eingesetzt. Vor jedem Schergang wurde das System durch zweimaliges Spülen mit 0,2 M HCl, dreimaliges Spülen mit 0,2 M NaOH und fünfmaliges Spülen mit dest. H₂O gereinigt. Nach dem Aufziehen der Probe wurden die Luftblasen in der Flüssigkeitssäule entfernt. Für den Erhalt der kleinen Fragmente (1,3-1,5 kbp) wurde die Geschwindigkeitsstufe 4 und für den Erhalt der großen Fragmente (2,5-3 kbp) die Geschwindigkeitsstufe 11 gewählt. Für die vollständige Fragmentierung der DNA wurden jeweils 25 Zyklen durchgeführt. Anschließend wurde die Probe aus dem System eluiert und die DNA-Fragmente mittels PEG gefällt (Kapitel 5.4.6) und in 30 µl H₂O wieder aufgenommen. Im Anschluss folgte für die Ligation (Kapitel 5.6.5) die Herstellung glatter DNA-Enden (Kapitel 5.6.3.2).

4.6.4 Herstellung glatter DNA-Enden

4.6.4.1 Variante mit T4-DNA-Polymerase

T4-Puffer:	67 mM Tris (pH 8,0), 6,7 mM MgCl ₂ , je 25 µM dNTP's, 10mM β-Mercaptoethanol
dNTP-Mix :	je 1mM dNTP

Die Eliminierung überstehender 3'-Enden und das Auffüllen der 5'-Enden von DNA-Fragmenten erfolgte modifiziert nach Denninger (1983) durch Behandlung mit T4-DNA-Polymerase und Desoxyribonukleotiden.

Zu diesem Zweck wurden die gefällten DNA-Fragmente in 25 µl T4-Puffer resuspendiert und anschließend wurden 3 µl T4-DNA-Polymerase zugegeben. Es folgte eine Inkubation bei

15°C für 2 h. Nach Zugabe von 3 µl dNTP-Mix wurde für weitere 30 min bei 15°C inkubiert. Die Termination der Reaktion erfolgte durch Kühlung auf Eis. Die Entfernung des Enzyms erfolgte durch Phenol-, Phenol/Chloroform-, Chloroform-Extraktion. Danach wurde zur Entsalzung der Probe eine Fällung mit PEG (Kapitel 5.4.6) durchgeführt. Das Präzipitat wurde für die nachfolgende Ligation (Kapitel 5.6.5) in 6 µl H₂O aufgenommen.

4.6.4.2 Variante mit T4-DNA-Polymerase und Klenow-Fragment

10x T4-Puffer:	0,1 M Tris-HCl (pH 7,9), 0,1 M MgCl ₂ , 0,5 M NaCl, 10 mM DTT, steril filtriert
dNTP-Mix:	0,25 mM je dNTP

Bei dieser Variante wurde neben der T4-DNA-Polymerase zusätzlich Klenow-Fragment für die Herstellung glatter DNA-Enden eingesetzt (pers. Mitteilung Michael Kube, MPI Berlin). Die, aus der Fragmentierung erhaltenen, DNA-Fragmente wurden in 30 µl H₂O aufgenommen. Für die Auffüllreaktion wurden 4 µl T4-Puffer, 2 µl dNTP-Mix, 2 µl Klenow-Fragment (5U/µl) und 2 µl T4-DNA-Polymerase (3U/µl) zugesetzt und für 30 min bei RT inkubiert. Anschließend wurde Gelloading-Puffer zugesetzt und zur Trennung der DNA-Fragmente auf ein Agarosegel aufgetragen. Nach dem Ausschneiden der gewünschten Fragmentgröße wurde die DNA extrahiert (Kapitel 5.5.3) und in die Ligation (Kapitel 5.6.5) eingesetzt.

4.6.5 Dephosphorylierung

TAK-Puffer:	330 mM Tris-acetat (pH 7,9), 660 mM Kaliumacetat, 100mM Magnesium-acetat, 5 mM DTT
-------------	--

Um bei einer Ligation die Religation des Vektors zu verhindern, wurden die Enden der DNA mit alkalischer Phosphatase (calf intestine phosphatase = CIP) dephosphoryliert. Hierzu wurden 25 µg Vektor nach vollständiger Linearisierung gefällt und in 103 µl H₂O aufgenommen. Für den Reaktionsansatz wurden 12 µl TAK-Puffer und 5 µl CIP (1U / µl) zugegeben. Nach einer Inkubationszeit von 45-60 min bei 37°C erfolgte eine Hitzedenaturierung bei 75°C für 10 min. Die Entfernung der alkalischen Phosphatase erfolgte durch einen Enzymentferner-Kit (Micropure-EZ) von Amicon nach Anleitung. Anschließend wurde die DNA mittels PEG (Kapitel 5.4.6) gefällt und in H₂O aufgenommen. Nach der photometrischen Konzentrationsbestimmung wurde die Probe durch Zugabe von H₂O auf 80 fmol Vektor/µl eingestellt.

4.6.6 Ligation von DNA-Fragmenten

10x Ligase-Puffer: 0,3 M Tris / HCl, 10 mM DTT, 66,5 mM MgCl₂, pH 7,5

Für den Reaktionsansatz wurden zu 6 µl DNA-Fragmenten 1µl Vektor (80fmol), 1µl 10x Ligase-Puffer, 1 µl ATP (0,5 mM) und 1 µl T4-DNA-Ligase (1 U / µl) zugegeben. Die Inkubation erfolgte bei 16°C für 3-8 h oder alternativ bei 4°C über Nacht. Die Termination der Reaktion erfolgte durch Hitzedenaturierung bei 65°C für 10 min.

4.7 Einstellungsparameter für die vollautomatische Übertragung von Zellkolonien mit einem Picking- und Spotting-Roboter

Für die vollautomatische Übertragung von Zellkolonien in Mikrotiterplatten im 384er Format wurde der Picking- und Spotting-Roboter Modell „Q-Bot“ der Fa. Genetix eingesetzt. Standardmäßig wurde die Einstellung „blau-weiß-Selektion“ verwendet. Während des routinemäßigen Laboreinsatzes wurden die Parameter optimiert. Die Strategie bestand darin einerseits möglichst viele weiße Klone von der Agarplatte aufzunehmen und in die Mikrotiterplatte zu übertragen. Andererseits musste vermieden werden, dass sehr dicht beieinander bzw. ineinander gewachsene Klone gleichzeitig durch einen Pin aufgenommen und in ein Gefäß der Mikrotiterplatte übertragen werden.

Mit der eingesetzten Software können durch Einstellung der Parameter „Min. Roundness“ und „Min Axis Ratio“ die als Oval erscheinenden, ineinandergewachsenen Bakterienkolonien ausgeschlossen werden. Um sicherzustellen, dass nur Einzelkolonien vom System aufgenommen werden, wurden mit den Werten 0,8 für „Min. Roundness“ und 0,85 für „Min. Axis Ratio“ stringente Parameter gewählt. Mit dem Parameter „Proximity Detection“ wird die Auswahl einer Kolonie in Abhängigkeit des minimal nötigen Abstands zu Nachbarkolonien festgelegt. Hierfür wurde der Wert 6 verwendet.

Die vollständige Animpfung der Mikrotiterplatte konnte durch zweimaliges Eintauchen der Pins in das Kulturmedium bei einer jeweiligen Verweildauer von 1 Sekunde erzielt werden. Die vollständige Sterilisation der Pins wurde durch 3 Zyklen im Sterilisationsbad mit 80%igem Ethanol erreicht. Für die Trocknung der Pins mittels Heizluft im „Dryer“ erwiesen sich 10 Sekunden als ausreichend. Vor dem Start des nächsten Pickingzyklus wurde noch eine Abkühlungsphase von 8 Sekunden eingestellt. Unter Verwendung dieser Parameter und der Einbrechnung der jeweiligen Einrichtungszeit konnte eine Übertragungsgeschwindigkeit von 2100 – 2200 Kolonien pro Stunde erzielt werden.

4.8 Polymerase-Kettenreaktion (PCR)

Die Polymerase-Kettenreaktion (Saiki et al., 1988) ist eine der wichtigsten Standardmethoden der Molekularbiologie geworden. Diese *in vitro* Methode erlaubt durch enzymatische Synthese die milliardenfache Vervielfältigung einzelner DNA-Moleküle. Amplifiziert wird ausgehend von zwei einander entgegengerichteten Oligonukleotiden (Primer), die als Startmoleküle für eine thermostabile DNA-Polymerase aus *Thermus aquaticus* (Taq-Polymerase) dienen. Die Vervielfältigung wird durch die zyklische Wiederholung folgender Reaktion erreicht: Hitzedenaturierung der Ziel-DNA, Anlagerung der Primer und Synthese des komplementären Doppelstranges. Dieses führt zu einer exponentiellen Akkumulation der Ziel-DNA, deren Enden durch die 5' Enden der Primer definiert sind (Newton und Grainger, 1994).

Wenn die Fehlerquote in der Sequenz des amplifizierten Produkts sehr niedrig sein soll besteht zusätzlich die Möglichkeit thermostabile DNA-Polymerasen mit „proof reading“ Aktivität einzusetzen. Ein solches Korrekturleseenzym überprüft die von der Taq-Polymerase synthetisierten Sequenzen und verbessert eingebaute Fehler. Die Fehlerhäufigkeit kann dadurch um einen Faktor von ca. 6 vermindert werden. In der vorliegenden Arbeit wurde die „Vent-DNA-Polymerase“ (NEB) aus *Thermococcus litoralis* eingesetzt.

4.8.1 PCR zur DNA-Fragmentgewinnung für die Sequenzierung

10x Taq-Puffer:	125 mM Tris (pH 9,5), 50 mM MgCl ₂ , 150 mM (NH ₄) ₂ SO ₄
10x Vent-Puffer :	10 mM KCl, 10 mM (NH ₄) ₂ SO ₄ , 20 mM Tris-HCl (pH 8,8), 2 mM MgSO ₄ , 0,1% Triton X-100
dNTP-Mix :	2,5 mM je dNTP

Im üblichen Verfahren wird die zu sequenzierende DNA in Bakterienzellen durch deren Kultivierung vermehrt, anschließend mittels kommerzieller Präparationskits (Anionenaustauschersäulen) oder durch Fällungsverfahren isoliert und dann sequenziert.

Bei dieser hier vorgestellten Methode wurde das zu sequenzierende DNA-Bruchstück mittels geeigneter Primerwahl durch PCR vermehrt. Dazu wurden die Bakterienzellen aus den Mikrotiter-Platten (z.B. "shotgun"-Banken) mit einem Stempel in die PCR-Ansätze übertragen oder es wurden ca. 25 ng DNA eingesetzt. Die PCR wurde in 20 µl-Ansätzen mit 1U DNA-Polymerase (Taq-Polymerase oder Vent-Polymerase), 1,5 µl dNTP-Mix, 5 pmol je Primer und 2 µl des entsprechenden Puffers durchgeführt. Je nach verwendeter DNA-Polymerase wurden unterschiedliche PCR-Programme verwendet:

Variante A: Programme bei Verwendung von Bakterienzellen

Taq-Polymerase:

95°C – 4 min

95°C – 20s }

65°C – 2 min 10s }

72°C – 10 min

Vent-Polymerase:

95°C – 4 min

95°C – 20s }

57°C – 15s }

72°C – 5 min

Variante B: Programme bei Verwendung von DNA

Taq-Polymerase:

95°C – 2 min

95°C – 20s }

65°C – 2 min 10s }

72°C – 10 min

Vent-Polymerase:

95°C – 2 min

95°C – 20s }

57°C – 15s }

72°C – 5 min

Im Anschluss wurde die DNA zur Entfernung der dNTP's und der Primer gefällt. Dazu wurden 14 µl (0,7 Volumen) Isopropanol zugegeben. Nach einer Inkubationszeit von 20 min (RT) wurde bei 2500 x g und 15°C für 20 min zentrifugiert. Danach wurde die Mikrotiter-Platte zur Entfernung des Überstands mit einem Papiertuch bedeckt und auf dem Kopf bei 100 x g für 1 min zentrifugiert. Die Entsalzung erfolgte durch Zugabe von 20 µl 70%igem Ethanol und anschließender Zentrifugation bei 2500 x g und 15°C für 15 min. Nach der Entfernung des Überstands mit der gleichen Prozedur wie oben beschrieben wurde das Präzipitat für 5 min bei RT getrocknet und in 10 µl H₂O aufgenommen. Hiervon wurden für die Sequenzierung des Amplifikats von beiden Seiten je 4 µl eingesetzt. Die verbliebenen 2 µl konnten zur Prüfung auf Nebenprodukte und zur Mengenabschätzung auf ein Agarosegel aufgetragen werden.

4.8.2 Herstellung DIG-11-dUTP markierter Sonden für Hybridisierungsexperimente

10x *Taq*-Puffer: 125 mM Tris (pH 9,5), 50 mM MgCl₂, 150 mM (NH₄)₂SO₄

dNTP-Mix (DIG): je 2,5 mM dTTP, dGTP und dCTP,
 2,45 mM dATP, 50 nM DIG-11-dUTP

dNTP-Mix : 2,5 mM je dNTP

Die Herstellung von DIG-11-dUTP markierter Sonden erfolgte auf zwei unterschiedlichen Wegen. Einerseits wurden am 5'-Ende DIG-11-dUTP markierte Primer verwendet, andererseits wurde DIG-11-dUTP dem dNTP-Mix zugefügt. Die PCR wurde in 20 µl-Ansätzen mit 1U Taq-Polymerase, 1,5 µl dNTP-Mix, 5 pmol je Primer und 2 µl des entsprechenden Puffers durchgeführt. Es wurden entweder Bakterienzellen, in der die zu markierende DNA-Sequenz enthalten war oder es wurden ca. 25 ng DNA eingesetzt. Hierfür wurden unterschiedliche PCR-Programme verwendet :

Variante A: Programm bei Verwendung von Bakterienzellen

Taq-Polymerase:

95°C – 4 min
 95°C – 20s }
 } 40x
 65°C – 2 min 30s }
 72°C – 10 min

Variante B: Programm bei Verwendung von DNA

Taq-Polymerase:

95°C – 2 min
 95°C – 20s }
 } 30x
 65°C – 2 min 30s }
 72°C – 10 min

Anschließend wurden die PCR-Ansätze mit Gelloader-Puffer versetzt und auf ein Agarosegel aufgetragen. Die Hybridisierungssonden wurden aus dem Gel ausgeschnitten und in 1,5 ml Eppendorf-Gefäße überführt. Bis zum Einsatz in die Hybridisierung wurden diese bei –20°C gelagert.

4.8.3 Sequenzierung von DNA

Sowohl beim ABI als auch beim LI-COR erfolgte die enzymatische Sequenzierreaktion nach Sanger et al. (1977), die auf einem nukleotidspezifischem Kettenabbruch durch den Einbau eines Didesoxy-Nukleotids beruht. Das gängige Laborverfahren ist die PCR-Sequenzierung, bei der die PCR unter Kombination mit Didesoxynukleotiden zur Generierung einer geeigneten Sequenzierauftragslösung verwendet wird. Der zentrale Mechanismus der PCR nutzt

das Schmelzverhalten der DNA aus. Oberhalb einer bestimmten Temperatur trennen sich die beiden Stränge der DNA. Rasche Abkühlung bewirkt, dass die Einzelstränge kaum wieder renaturieren können. Die im Reaktionsgemisch vorhandenen Primer haben nun die Möglichkeit sich an ihre komplementären Stellen der DNA zu binden. Anschließend wird das Reaktionsgemisch auf das Temperaturoptimum der thermostabilen Polymerase (Taq aus *Thermus aquaticus* bzw. eine modifizierte T7 Polymerase (Koop *et al.* 1993) aufgeheizt und die DNA-Einzelstränge werden wieder zu Doppelsträngen ergänzt. Da im Reaktionsgemisch Didesoxynukleotide vorliegen erfolgt ein Abbruch der Polymerase-Strangregeneration an statistisch über den Matrizenstrang verteilten Stellen. Eine erneute Hitzedenaturierung der DNA beginnt einen zweiten Zyklus, usw. In der Folge entstehen so über die statistischen Strangabbrüche der Polymerasereaktion DNA-Fragmente, deren Länge sich um jeweils ein Nukleotid unterscheiden, was bei der anschließenden Elektrophorese zur Sequenzgewinnung genutzt wird.

Der Unterschied bei den Sequenzierverfahren liegt in der verwendeten Chemie. Bei der ABI-Sequenzierung wird die Reaktion in einem Probengefäße durchgeführt, da die Didesoxynukleotide A, C, G und T jeweils unterschiedlich gelabelt sind. Dadurch kann die Probe auf eine Spur des Sequenziergels aufgetragen werden, da bei Anregung jeweils sichtbares Licht unterschiedlicher Wellenlänge (Farbe) emittiert wird. Durch Detektion mit einer CCD-Farbkamera kann so die Basenabfolge ermittelt werden.

Beim LI-COR sind die Primer am 5'-Ende mit IRD700 oder IRD800 gelabelt. Die Reaktion wird in 4 Probengefäßen durchgeführt in denen jeweils nur von einem Desoxynukleotid (A, C, G, T) zusätzlich die Didesoxynukleotide vorliegen. Die 4 Proben werden entsprechend auf vier Spuren des Sequenziergels in vorgegebener Reihenfolge aufgetragen, wodurch eine Ermittlung der Basenabfolge ermöglicht wird.

Für die beiden Systeme wurden für die Sequenzierung von "shotgun"-Banken (Millipore DNA-Präparation im 96er-Raster; Plasmide) unterschiedliche Standard-PCR-Protokolle verwendet. Zusätzlich wurden für die Sequenzierung von PCR-Produkten oder von unfragmentierter Ausgangs-DNA (BAC-, PAC-, Cosmid-DNA) vom Standard abweichende PCR-Protokolle eingesetzt.

ABI-Protokolle

Natriumacetat-Puffer: 1,5 M Natriumacetat, 250 mM EDTA, pH > 8,0

Gellade-Puffer :

je nach verwendeter DNA unterschiedliche Mengen eingesetzt. Von PCR-Produkten wurden 120-200 ng, von Plasmid-DNA 0,6-0,8 µg und von BAC-DNA 1-2 µg verwendet. Für PCR-Produkte und Plasmid-DNA wurden als weitere Komponenten 4 µl Sequenziermix und 1 µl Sequenzier-Primer (5 pmol / µl) zugegeben. Der Ansatz wurde mit bidest. H₂O auf 10 µl aufgefüllt. Für den Sequenzieransatz von BACs wurden 1 µl DNA, 1 µl Sequenzier-Primer (5 pmol / µl) und 8 µl Sequenzier-Mix eingesetzt.

Standartsequenzier-PCR

95°C – 15 s	}	} 15x
55°C – 15 s	}	
70°C – 1 min	}	

95°C – 15 s	}	} 15x
70°C – 1 min	}	

Sequenzier-PCR für PCR-Produkte

95°C – 3 min

95°C – 15 s	}	} 37x
55°C – 15 s	}	
70°C – 1 min	}	

Sequenzier-PCR für BAC-DNA

96°C – 2 min

96°C – 10s	}	} 55x
55°C – 10s	}	
60°C – 4 min	}	

Für die anschließende Fällung wurden 2 µl Natriumacetat-Puffer und 32 µl 100% Ethanol zugegeben, durchmischt und für 15 min bei RT inkubiert. Die Zentrifugation wurde in einer Mikrotiterplatten-Zentrifuge bei 15°C und 2500 x g für 40 min durchgeführt. Zur Entfernung des Überstands wurde ein Papiertuch auf die Mikrotiterplatte gelegt und auf dem Kopf stehend für 1 min bei 100 x g zentrifugiert. Nach dem Trocknen bei RT für 5 min wurde das Präzipitat in 3 µl Gellade-Puffer aufgenommen. Vor dem Auftrag auf das Gel werden die Proben für 3 min bei 95°C denaturiert und dann sofort auf Eis gekühlt.

LI-COR-Protokolle

Formamid-Loading-Puffer:

Für den Sequenzieransatz wurden von Plasmid-DNA 1 µg und von BAC-DNA 2 µg eingesetzt. Als weitere Komponenten wurden 1,5 pmol von dem IRD700 gelabelten Primer und/oder 3 pmol von dem IRD800 gelabelten Primer eingesetzt und mit bidest. H₂O auf ein

Volumen von 20 µl aufgefüllt. Dieser Ansatz wurden auf 4 Probengefäße gleichmäßig verteilt und 1,5 µl A-, C-, G- bzw. T-Mix zugeben.

Standartsequenzier-PCR

95°C – 3 min

95°C – 15 s	}	
55°C – 30 s		} 50x
68°C – 30 s		

Sequenzier-PCR für BAC-DNA

96°C – 3 min

96°C – 15 s	}	
55°C – 30 s		} 50x
68°C – 30 s		

Danach wurden die Proben für 5 min bei 90°C eingeeengt und 2 µl Formamid-Loading-Puffer zugegeben. Vor dem Auftrag auf das Gel wurden die Proben für 2 min bei 70°C denaturiert und anschließend sofort auf Eis gekühlt. Es wurden je 0,7 ml auf das Gel aufgetragen.

4.9 Gelelektrophorese

4.9.1 Agarosegele

20 x TBE-Puffer: 216 g Tris, 106,8 g Borsäure, 14,8 g EDTA mit bidest. H₂O auf 1 L Gesamtvolumen aufgefüllt.

Ethidiumbromid-Lösung: 5 mg/ml in H₂O

6 x Gelladepuffer: 0,2 % (w/v) Bromphenolblau, 0,2 % (w/v) Xylencyanol, 30 % (w/v) Glycerin.

Zur Anfertigung eines 0,8%igen Standardgels wurden 0,8 g Agarose pro 100ml 1 x TBE-Puffer eingesetzt und in einem Mikrowellenherd so lange gekocht, bis die Agarose vollständig gelöst war. Dann wurde die flüssige Agarose mit 5 µl Ethidiumbromid vermischt und nach einer Abkühlphase in eine Gelkammer gegossen. Zur Formung der Geltaschen wurde ein Kamm eingesetzt. Nach dem Erstarren wurde das Gel in eine mit 1 x TBE-Puffer befüllte Elektrophoresekammer gesetzt. Die Proben wurden vor dem Auftragen mit 1/6 Volumen des 6 x Gelladepuffers versetzt und auf das Gel aufgetragen. Für die Erkennung der Größe eines Fragments wurden Größenstandards parallel mit aufgetragen. Die Elektrophoresebedingungen richteten sich nach dem jeweiligen Einsatzzweck.

Für die Auftrennung der DNA-Fragmente für die Herstellung von "shotgun"-Banken startete die Elektrophorese mit 130 V für 12 min und wurde dann bei 100 V für weitere 2 h fortgeführt.

Im Rahmen der Sequenzierung wurden analytische Restriktionen eingesetzt um durch Vergleich des Verdaumusters der präparierten DNA mit dem Verdaumuster, das dem Klon bei seiner Versendung beigelegt wurde, die Korrektheit des Klons zu überprüfen. Zu diesem Zweck wurde die Elektrophorese bei 60 V für 16 h durchgeführt.

In allen anderen Fällen wurden Spannungen von 4-6 V pro cm Elektrodenabstand angelegt und die Elektrophorese für 2-5 h durchgeführt.

Die Geldokumentation erfolgte am „Fluor-STM MultiImager“ von BIO-RAD. Zur präparativen Gewinnung von DNA-Fragmente wurden die Banden unter UV-Licht (360 nm) aus dem Agarosegel ausgeschnitten.

4.9.2 Polyacrylamidgele für die Sequenzierung

TEMED

10% APS-Lösung: 1.0 g Ammoniumpersulfat (APS) mit H₂O bidest. auf 10 ml auffüllen.

10 x TBE: 108 g Tris-Base, 55 g Borsäure, 7,4 g EDTA mit bidest. H₂O auf 1 L Gesamtvolumen aufgefüllt.

Acrylamid-Gele für ABI 377:

48cm WTR (90 B/h = 2X RUN)

4.25 %iges Acrylamidgel :

21.0g Harnstoff

7.1ml 30% Acrylamid

6.0ml 10 x TBE-Puffer

21.0ml H₂O bidest.

Polymerisationsstart durch Zugabe von:

350 µl 10%iges APS

15 µl TEMED

Acrylamid-Gele für LI-COR:

66 cm x 0,2 mm

4,5%iges Acrylamidgel

21 g Harnstoff

5,4 ml 50% Long-Ranger-Gel-Solution

7,2 ml 10 x TBE

27,5 ml H₂O bidest.

Polymerisationsstart durch Zugabe von:

400 µl 10%iges APS

40 µl TEMED

Die Glasplatten wurden vor dem Lauf mit einem Detergenz (Alconox) behandelt und mit destilliertem Wasser gespült. Mit einem (9:1) Gemisch aus Isopropanol und H₂O wurden die Platten nachbehandelt. Auf die getrockneten Glasflächen wurden dann die Spacer aufgesetzt und die Gelkassette zusammengebaut.

Die Gellösung wurde je nach verwendetem System angesetzt. Nachdem der Harnstoff vollständig gelöst ist wurde die Lösung über ein 0,2 µm Filter geklärt. Das Filtrat wurde vor dem Einsatz für ca. 5 min entgast. TEMED und 10%ige APS-Lösung wurden in den entsprechenden Mengen zum Schluss zugegeben, wodurch die Polymerisation startete. Die Flüssigkeit wurde nun zügig zwischen die Glasplatten der Gelkassette gegossen, wobei darauf zu achten war, dass sich keine Luftblasen im Gel befanden. Nach Einfüllen der Polymerisationslösung wurde sogleich die Rückseite des Haifischzahn-Kamms eingesetzt. Dadurch wurde eine Geltasche erzeugt, die fast die ganze Gelbreite überspannte. Die Polymerisation zum Gel erfolgte innerhalb der nächsten 2 Stunden. Während der Polymerisation sollte die Kassette horizontal lagern. Anschließend wurden die Spitzen des Haifischzahn-Kamms leicht in das Gel am Taschenboden eingedrückt um so die Auftragungs-taschen zu erzeugen.

4.9.2.1 Sequenzierlauf

Vor dem Auftrag der Proben auf das Sequenziergel wurden diese für 3 min bei 95°C denaturiert und anschließend sofort auf Eis gekühlt. Als Laufpuffer wurden je nach System unterschiedliche Verdünnungen des 10 x TBE-Puffers eingesetzt. Für den ABI wurde 1 x TBE-Puffer und für den LI-COR 0,8 x TBE-Puffer verwendet.

Beim ABI wurden für das Gelsystem 48 cm x 0,2 mm WTR (90 B/H = 2X RUN) folgende Laufbedingungen gewählt:

3000 V, 50 mA, 48 W, 50°C Geltemperatur, 40 mW Laserleistung bei einer Dauer des Laufs von 10 h.

Beim LI-COR wurden für das Gelsystem 66 cm x 0,2 mm folgende Laufbedingungen gewählt:

2800 V, 40 mA, 50 W, 45°C Geltemperatur bei einer Dauer des Laufs von ca. 16h.

4.10 Zusammenfügen von „shotgun“-Sequenzdaten mittels eines Assemblierungsprogramms

Die, aus der Sequenzierung der Klone einer „shotgun“-Bank erhaltenen, Rohsequenzen wurden mit dem Programm „Pregap4“ aufgearbeitet. Dieses Programm legte im ersten Schritt ein Projektverzeichnis an. Anschließend wurden die Dateien in ein „scf“-Format konvertiert, das für die Assemblierung der Rohsequenzen mit dem Programm „Gap4“ notwendig war und im Projektverzeichnis gespeichert. Ferner wurde mit dem Unterprogramm „Phred“ jeder Base einer Rohsequenz anhand des Elektropherogramms ein Qualitätswert zugeordnet. In der Folge wurden alle Sequenzen oder Sequenzteile durch Sequenzvergleiche identifiziert und entfernt bzw. maskiert, die nicht zum Sequenzbruchstück des Chromosoms gehören. Die Bereiche am Anfang der Rohsequenz, die zum verwendeten Sequenzierungsvektor gehören, wurden mittels der Funktion „sequence vector clip“ abgeschnitten. Vollständig mit dem Sequenzierungsvektor übereinstimmende Rohsequenzen, also von Klonen, die keine Insertion im Vektor tragen, wurden ebenfalls entfernt. Unter der Funktion „Blast“ wurden die Rohsequenzen mit der Sequenz des *E. coli*-Genoms verglichen und solche entfernt, die zu mehr als 90% mit der genomischen Sequenz übereinstimmten. Die Rohsequenzen bzw. Teile davon, die mit dem Klonierungsvektor übereinstimmen, wurden mit der Funktion „Crossmatch“ identifiziert und maskiert.

Im Anschluss wurden die so vorbereiteten und im Projektverzeichnis gespeicherten Rohsequenzen mit „Gap4“, einem Programmpaket zur Assemblierung von Sequenzen, weiter bearbeitet. Dazu wurde das Unterprogramm „Phrap assembly“ zum Zusammenfügen der Rohsequenzen verwendet. Dieses Programm arbeitete in zwei Schritten. Im ersten Schritt wurden die Rohsequenzen in ihrer Gesamtqualität beurteilt. Diese wurde aus den Qualitätswerten ermittelt, die zuvor mit dem Unterprogramm „Phred“ jeder einzelnen Base dieser Sequenz zugeteilt wurde. Mit zunehmender Leseweite wird die Qualität der Sequenz schlechter. Wird die, vom Programm vorgegebene Mindestqualität unterschritten, so wird der nachfolgende Bereich maskiert. Ebenso wurde mit dem Sequenzanfang verfahren. Ist die Rohsequenz durchgehend schlecht, so wurde sie nicht für die Assemblierung herangezogen. Letztendlich erfolgte im zweiten Schritt das automatische Zusammenfügen anhand der Überlappungsbereiche der verbliebenen „shotgun“-Sequenzen.

Im Normalfall weist die so erhaltene Sequenz immer noch Lücken auf, die durch Ausfälle bei der Sequenzierung oder durch Klonierungslücken bei der Herstellung von „shotgun“-Banken entstehen. Die lückenfrei zusammengefügt Sequenzen bezeichnet man als Contigs. Das Schließen der Lücken zwischen den Contigs erfolgte wieder in mehreren Schritten. Zuerst

wurden die Enden der Contigs manuell editiert, um so vom Auswertungsprogramm falsch erkannte Basen zu korrigieren. Dann wurde versucht mittels der Funktion „find internal joins“ des Programms „Gap4“ einzelne Contig manuell zusammenzufügen. Anschließend wurden die noch vorhandenen Sequenzlücken zwischen den Contigs durch primer walking auf ausgesuchten Klonen der „shotgun“-Bank oder direkt auf der entsprechenden PAC- bzw. BAC-DNA geschlossen. Ebenso wurde mit Sequenzbereichen verfahren, die noch nicht den Qualitätsstandards entsprechen, die im Rahmen des Human-Genom-Projekts gefordert sind. Dieser ist erreicht, wenn die gesamte, zusammengefügte Sequenz „doppelsträngig“ vorliegt und jede Basen mindestens dreifach abgedeckt ist. Die erhaltene Sequenz wird dann als „doppelsträngig“ bezeichnet wenn sie aus „shotgun“-Sequenzen zusammengefügt wurde, die in entgegengesetzter Richtung sequenziert wurden, oder wenn alle „shotgun“-Sequenzen in gleicher Richtung liegen, die Sequenzierung aber mit unterschiedlicher Chemie erfolgte (z. B. ABI und LI-COR).

4.11 Hybridisierungsexperimente

4.11.1 Herstellung der Filter für Hybridisierungen

Denaturierungspuffer:	468 mM NaOH, 1,4 M NaCl in bidest. H ₂ O gelöst, Gesamtvolumen 1 L.
Tris-Puffer:	160 mM Tris-Base, 840 mM Tris-HCl, 1,5 M NaCl in bidest. H ₂ O, Gesamtvolumen 1 L.
Pronase-Puffer:	28 mM Tris-Base, 22 mM Tris-HCl, 100 mM NaCl, 50 mM EDTA, 31,5g Sarcosyl NL-30, 250 mg Pronase in bidest. H ₂ O gelöst, Gesamtvolumen 1 L.

Die Herstellung von Filtern für Hybridisierungen erfolgte nach dem Verfahren von Hoheisel *et al.* (1991), das teilweise mit dem Verfahren von „Roche Molecular Biochemicals“ kombiniert wurde.

Die Übertragung der Bakterienkolonien aus den 384er-Mikrotiter-Dauerkulturen auf den Filter erfolgte mit dem Picking- und Spotting-Roboter „Q-Bot“ der Fa. Genetix. Die Sterilisation des Innenraums vom Roboter erfolgte durch UV-Licht. Die Fixierung der Filter auf den dafür vorgesehenen Platten wurde durch das Unterlegen von mit 2YT-Medium befeuchtetem 3MM-Whatmanpapier erreicht. Dazu wurde das auf Plattengröße zugeschnittene Whatmanpapier in einer Schale mit 2YT-Medium durchweicht. Bevor diese auf die Platten gelegt wurden ließ

man das überschüssige Medium gut abtropfen. Anschließend wurden die Filter von der Unterseite durch Auflegen auf die Flüssigkeitsoberfläche des Mediums befeuchtet und nach Abtropfen des überschüssigen Mediums auf das Whatmanpapier gelegt. Durch diese Vorgehensweise konnte ein Verlaufen der aufgespotteten Kolonien verhindert werden.

Ein Filter von 22,2 x 22,2 cm ist jeweils in 6 Felder unterteilt. Das Aufspotten der Kolonien erfolgte je nach Anzahl der 384er-Mikrotiter-Dauerkulturplatten in einem 3 x 3- bzw. 4 x 4-Raster. Jeder Klon einer Platte wurde in Abhängigkeit der Plattennummerierung zweimal in einem bestimmten Muster auf den Filter übertragen. Dadurch konnte bei einem positiven Signal neben der Lage des Klons innerhalb einer Mikrotiterplatte auch die entsprechende Platte, in der sich der Klon befindet, identifiziert werden. Durch diese doppelte Auftragung konnten, je nach Raster, maximal 4 bzw. 8 Mikrotiterplatten auf ein Feld gespottet werden. Alle 6 Felder waren in diesem Fall identische Kopien, um ausreichend Filter für die Hybridisierungen zur Verfügung zu haben. Nach der Übertragung der Kolonien auf den Filter wurde dieser auf eine 2YT-Agarplatte gelegt. Dabei war darauf zu achten, dass keine Luftblasen zwischen Filter und Agar waren um zu verhindern, dass einzelne Kolonien nicht anwuchsen. Es folgte eine Inkubation bei 37°C über Nacht.

Anschließend wurde der Filter auf mit Denaturierungspuffer befeuchtetes 3MM-Whatmanpapier gelegt und für 4 min bei RT inkubiert. Dieser Vorgang wurde nochmals wiederholt. Dann wurde der Filter auf mit Tris-Puffer gesättigtes 3MM-Whatmanpapier gelegt und für mindestens 5 min bei RT inkubiert. Danach wurde der Filter vorsichtig in 400 ml auf 37°C vorgewärmten Pronase-Puffer gelegt und bei 37°C ohne schütteln für 1 h inkubiert und dann auf trockenes 3MM-Whatmanpapier gelegt und über Nacht bei RT getrocknet. Zur DNA-Fixierung wurden die Filter für 5 min mit UV-Licht bestrahlt und anschließend in sechstel zerschnitten. Die Lagerung erfolgte bei RT gelagert.

Alternativ wurden die Filter zur Entfernung der Zelltrümmer (Roche) direkt nach der Pronase-Inkubation für 5 min zur DNA-Fixierung mit UV-Licht bestrahlt. Anschließend wurde der Filter zwischen zwei mit bidest. H₂O befeuchtete 3MM-Whatmanpapiere gelegt und gewalzt. Das obere 3MM-Whatmanpapier wurde vorsichtig vom Filter abgezogen und der Vorgang mehrmals wiederholt, bis keine Zellreste mehr erkennbar waren. Nachdem der Filter über Nacht auf trockenem 3MM-Whatmanpapier getrocknet war wurde dieser in sechstel zerschnitten und bei RT gelagert.

4.11.2 Hybridisierung

modifiziert nach Hoheisel *et al.* (1991) und E. Maier (1994)

Church-Puffer:	0,5M Na-Phosphat (pH 7,2), 7% (w/w) SDS, 1 mM EDTA (pH 8,0), 0,1 mg/ml Fischsperma-DNA
TN-Puffer:	0,1 M Tris-HCl, 0,15 M NaCl, pH 7,5
TNB-Puffer:	0,1 M Tris-HCl, 0,15 M NaCl, 1% (w/v) blocking agent (Roche), pH 7,5
Anti-DIG-AP-Lösung:	0,15 U/ml Anti-DIG-AP in TNB-Puffer
TNM-Puffer:	0,1 M Tris, 0,1 M NaCl, 1 mM MgCl ₂ , pH 9,0
DM-Puffer:	0,1 M Diethanolamin, 1 mM MgCl ₂ , pH 9,0
Waschpuffer:	40 mM Na-Phosphat, 0,1% (w/w) SDS, pH 7,2
10 x blocking agent:	

Vor der Benutzung neuer und trocken gelagerter Filter werden diese in 200 ml Church-Puffer für mindestens 2 h bei 65°C inkubiert und anschließend bei 37°C in dem Puffer gelagert.

Die mit Dig-11-dUTP markierten Sonden (Kapitel 5.7.2) wurden einzeln oder als Gemisch für die Hybridisierung eingesetzt. Zum Lösen der Agarose, in der sich die Sonden befinden und zur Denaturierung der DNA wurden diese mit 50-100 µl H₂O versetzt und für 5 min in kochendem Wasser erhitzt. Nach Zugabe von 1-1,5 ml Church-Puffer wurde die Lösung gleichmäßig auf den auf Klarsichtfolie liegendem Filter verteilt und der Deckel der Klarsichtfolie aufgelegt. Eventuelle Luftblasen wurden vorsichtig von dem Filter durch Wischen entfernt. Anschließend wurden die Filter luftdicht in Haushaltsfolie verpackt und bei 65°C über Nacht inkubiert. Zur Entfernung unspezifisch gebundener Sonden wurden die Filter kurz vorgewaschen und dann in 500 ml Waschpuffer für 15 min bei 65°C inkubiert.

Anschließend folgte die Farbreaktion. Bei den hierfür notwendigen Waschschritten vor und nach der Antikörper-Reaktion wurde jeder Filter einzeln in entsprechenden Schalen mit 50 ml Puffer für 5 min bei RT unter Schwenken inkubiert. Vor jeder Überführung in den nachfolgenden Puffer ließ man den Filter gut abtropfen. Die Antikörper-Reaktion erfolgte in 30 ml Anti-DIG-AP-Lösung für 45 min bei RT unter Schwenken.

Folgende Schritte wurden durchgeführt:

TN-Puffer, pH 7,5

TNB-Puffer, pH 7,5

Anti-DIG-AP-Lösung, pH 7,5

TN-Puffer, pH 7,5

TNM-Puffer, pH 9,0

DM-Puffer, pH 9,0

Danach ließ man die Filter nacheinander auf 3MM Whatmanpapier kurz antrocknen. Anschließend wurden die Filter auf Klarsichtfolie gelegt und mittels eines Zerstäubers gleichmäßig mit wenig AttoPhos-Lösung besprüht. Nach Auflegen des Deckels der Klarsichtfolie auf den Filter wurde die AttoPhos-Lösung durch Wischen gleichmäßig verteilt. Für die Farbentwicklung wurden die Filter für 1-3 h im Dunkeln inkubiert. Die entwickelten Filter wurden unter UV-Licht unter Verwendung eines 520 nm Longpass-„cut off“-Filters photographiert und ausgewertet.

4.12 Durchführung der Annotierung

Im ersten Schritt wurden die Sequenzen einer sogenannten „first-pass“ Annotierung unterzogen. Diese erfolgte automatisch unter Verwendung des Programms „annomitter“ (Hornischer K. & Blöcker H., GBF). Dieses Programm vergleicht eine Sequenz mit unterschiedlichen Datenbanken, um Sequenzwiederholungen unterschiedlichen Typs bzw. ESTs zu finden. Ferner werden automatisch verschiedene Programme verwendet, mit denen auf der Sequenz codierende Exons vorhergesagt werden können, bzw. nach „Tandem Repeats“, „Inverted Repeats“, „Micro Satellites“ und „CpG Islands“ gesucht werden kann. Die erhaltenen Ergebnisse werden vom Programm zusammengefasst und als ein Ergebnis wiedergegeben. Auf der ersten Seite sind alle Programme und Datenbanken, die für die Erzeugung des Annotierungsergebnisses verwendet wurden, aufgelistet.

Die weitere Bearbeitung dieser „first-pass“-Annotierungsdaten erfolgte nach den Kriterien, die vom „Chromosom 21 Konsortium“ für die Vorhersage von Genen vorerst festgelegt worden sind (The chromosome 21 mapping and sequencing consortium, 2000). Diese Kriterien sind bewusst stringent gewählt worden, um eine größere Anzahl von falsch-positiv vorhergesagten Genen zu vermeiden. Diese Kriterien lauten:

- 1) Die Vorhersage möglicher codierender Exons ist auf die Anwendung der entsprechenden Programme „GRAIL“, „GENSCAN“ und „MZEF“ beschränkt. Ein Exon gilt als bestätigt, wenn es von zwei dieser Programme gefunden wurde.

- 2) ESTs sollen nur dann als Hinweis für eine mögliches Gen herangezogen werden, wenn die Sequenz des ESTs über den gesamten Bereich eine Homologie von mindestens 95% mit der genomischen DNA aufweist und auf dieser in zwei oder mehr Exons aufgeteilt vorliegt.
- 3) Ähnlichkeiten von Aminosäuresequenzen zu bekannten Proteinen oder funktioneller Domänen sollen nur dann als signifikanter Hinweis in betracht gezogen werden, wenn die Homologie über den gesamten Bereich mindestens 25%. Ferner soll die betrachtete Sequenz mindestens eine Länge von 50 Aminosäuren aufweisen.

Diese Kriterien wurden auf die Ergebnisse der „first-pass“-Annotierung angewendet, um aus dieser die signifikanten Daten herauszufiltern. Für die einfachere Bearbeitung wurden die textbasierten „first-pass“-Annotierungsdaten mit dem Programm „showseq“ (Hornischer K., GBF) in graphischer Form dargestellt. Im Anhang D (Kapitel 7.4.1) findet sich ein Beispiel für die graphische Darstellung der „first-pass“-Annotierungsdaten, das mit diesem Programm erhalten wurde. Die Bearbeitung der graphischen „first-pass“-Annotierung erfolgte in mehreren Schritten. Im ersten Schritt wurden alle Exons markiert, die von mindestens zwei der drei genannten Exon-Vorhersageprogrammen gefunden wurden. Allerdings wurde hier noch eine zusätzliche Einschränkung angewendet. Unterschieden sich die, an einer Stelle von zwei Programmen gefundenen Exons, deutlich in der Länge, so wurden diese nicht für die weitere Auswertung herangezogen. Im zweiten Schritt wurden alle aufgeführten ESTs anhand ihrer „Accession-Number“, dahingehend untersucht, ob diese in den textbasierten Annotierungsdaten mehrfach zu finden ist, also ein EST auf der genomischen DNA in mehrere Exons aufgeteilt ist. Diese wurden markiert und anschließend die Sequenzhomologie dieser ESTs mit der genomischen DNA überprüft. Ferner wurde überprüft, ob sich die Sequenz eines ESTs lückenfrei aus den Exons auf der genomischen DNA zusammenfügen ließ, die mittels dieser ESTs dort lokalisiert worden sind. Für die weitere Bearbeitung wurden nur die ESTs verwendet, die entsprechend den Vorgaben, eine Homologie von mindestens 95% aufwiesen und deren Sequenz sich durch Zusammenfügen der so gefundenen Exons lückenfrei erhalten lies. Die so erhaltenen Ergebnisse aus dem ersten und zweiten Schritt wurden zusammengefügt. Diese sind oftmals deckungsgleich. In einigen Fällen konnten Exons, die nur von einem der Exon-Vorhersageprogrammen gefunden wurden, anhand der ESTs bestätigt werden. Bei den bestätigten Exons wurden dann die Intron-Exon-Grenzen (Knippers R., 1985) überprüft und nötigenfalls korrigiert.

Um den Anfang eines möglichen Gens lokalisieren zu können wurden die, in den Annotierungsdaten aufgeführten, „CpG-Islands“ herangezogen, da diese in eukaryotischen Genomen oftmals am 5`-Ende, also am Anfang eines Gens zu finden sind (Gardiner-Garden M. *et al.*, 1987; Bird A. P., 1987). Dies wurde aber nur als Anhaltspunkt herangezogen.

Die Nukleotidsequenz wurde dann aus den Exons des möglichen Gens zusammengefügt. Danach wurde die so erhaltene Nukleotidsequenz ausgehend von den drei möglichen Leserastern in eine Aminosäuresequenz übersetzt (The Protein Machine, <http://www.ebi.ac.uk/translate/>). Für die weitere Bearbeitung wurde das Leseraster verwendet, mit dem sich eine Aminosäuresequenz erhalten ließ, die von einer möglichst geringen Anzahl oder überhaupt nicht von Stopcodons unterbrochen war. Wurde eine durchgehende Aminosäuresequenz erhalten, bei der sich kein Startcodon definieren ließ und die auch kein Stop-codon enthielt, so wurde die gesamte Sequenz für den Vergleich mit den Proteindatenbanken verwendet, um entsprechende Ähnlichkeiten mit bekannten Proteinen oder funktionellen Domänen zu analysieren. Für diesen Vergleich wurden die Datenbanken „SMART - Simple Modular Architecture Research Tool“ (<http://smart.embl-heidelberg.de/>) und „pfam-Protein families database“ (<http://www.sanger.ac.uk/Pfam/>) verwendet. Befand sich im Anfangsbereich der erhaltenen Aminosäuresequenz ein Stopcodon, so wurde die Sequenz ausgehend vom nächst folgenden Methionin (Startcodon) für den Vergleich herangezogen. Der Bereich vor dem Startcodon wurde als 5`-UTR (Untranslated Region) bezeichnet. Ebenso wurde verfahren wenn sich hinter dem Stopcodon noch Sequenz befand. Dieser Bereich wurde als 3`-UTR bezeichnet. In dieser Region wurde dann entsprechend nach dem Poly-A-Signal gesucht (DNA-Sequenz: AATAAA).

5 Zusammenfassung

Die Erbinformation des Menschen wird auf ca. 10^9 DNA-Nukleotiden codiert. Wie in jedem eukaryotischen Organismus wird diese genetische Information in Chromosomen organisiert. Der Mensch weist insgesamt 23 Chromosomenpaare auf. Die Basenabfolge des menschlichen Genoms wird im Human-Genom-Projekt analysiert. In der vorliegenden Arbeit werden Strategien vorgestellt, die zu Optimierungen in der systematischen Analyse genomischer DNA geführt haben. Zur Validierung der Methoden wurde DNA-Material des Chromosoms 9 und 21 verwendet.

Ein Ergebnis der vorliegenden Arbeit besteht in der Optimierung der Herstellung von "shotgun"-Banken. Durch Optimierung der Protokolle zur DNA-Präparation konnte der Anteil genomischer DNA der *E. coli*-Zellen an der Gesamtpräparationsmenge von 13,6% auf einen durchschnittlichen Wert von nur 3,1% verringert werden. Im Vergleich zum Ausgangsprotokoll kann mit dem verbesserten Protokoll die sechsfache Anzahl an „shotgun“-Banken im gleichen Zeitraum hergestellt werden. In der Summe der Optimierungen wurde ein sehr zuverlässiges Herstellungsverfahren für „shotgun“-Banken etabliert und in den Routinebetrieb übernommen.

Beim "hierachical shotgun"-Verfahren geht der verwendete Klonierungsvektor durch Fragmentierung der präparierten DNA mit in die „shotgun“-Bank ein. Durchschnittlich liegt der Anteil an Bakterienklonen, die ausschließlich Klonierungsvektor tragen, in einer „shotgun“-Bank bei ca. 10%. Zur Entfernung von Klonierungsvektor tragenden Bakterienklonen aus "shotgun"-Banken wurde basierend auf Hybridisierungsexperimenten ein geeignetes Verfahren etabliert und bei zwei Banken erfolgreich eingesetzt. In beiden Fällen konnte der unerwünschte Anteil von Bakterienklonen vollständig entfernt werden. Zugleich wurden auch die nicht angewachsenen Klone aus der Bank entfernt.

Standardmäßig wird in die Sequenzierreaktion Plasmid-DNA eingesetzt, die zuvor durch Verwendung von Plasmid-DNA-Präparationskits erhalten wurde. Diese sind auf die parallele Präparation von 96 Proben beschränkt und verursachen einen erheblichen Kostenaufwand. Heutige Sequenzierautomaten ermöglichen einen sehr hohen Probendurchsatz. Durch die Formatbeschränkung dieser Präparationskits zeichnet sich ein Engpass ab, der nur durch die Anschaffung kostspieliger Präparationsroboter zu beheben ist. Eine kostengünstige Alternative bietet die Template-DNA-Gewinnung durch PCR-Amplifikation. Hierzu wurden Bakterienzellen direkt in den PCR-Ansatz gegeben, durch Hitzedenaturierung der Zellen die Template-DNA freigelegt und anschließend amplifiziert. Durch Einführung und Optimierung eines

Fällungsschritts konnten die PCR-Produkte so weit aufgereinigt werden, das sie problemlos in die Sequenzierreaktion eingesetzt werden können. Im Vergleich zum herkömmlichen Sequenzierungsverfahren mit isolierter Plasmid-DNA zeigte sich kein Qualitätsverlust bei den so erhaltenen Sequenzierergebnissen. Ferner können mit diesem Verfahren 384 Proben parallel bearbeitet werden. Dadurch wurde neben der deutlichen Materialkostenreduktion auch der Arbeitsaufwand verringert.

Zusammengenommen können durch diese Optimierungen die vorhandenen Sequenzierkapazitäten sehr viel ökonomischer ausgenutzt werden.

Ferner wurden im Rahmen dieser Arbeit zwei Klone von Chromosom 21 vollständig sequenziert. Anschließend wurde mit Hilfe eines Programmpakets, das die Sequenz mit verschiedenen Datenbanken vergleicht und Funktionsvorhersagen erstellt, annotiert. Durch Auswertung dieser Annotationsdaten nach festgelegten Kriterien konnten 2 Gene gefunden werden.

Bei der Herstellung von "shotgun"-Banken ist es notwendig, sehr viele Klone von Agarplatten aufzunehmen und in Mikrotiterplatten im 384er Format zu übertragen. Zur Erleichterung dieser Tätigkeit wurde ein mechanisches Picking-Tool entwickelt und zum Patent angemeldet. Hiermit können annähernd Übertragungsgeschwindigkeiten erzielt werden, wie sie sonst nur von teuren Robotern erreicht werden.

6 Literaturverzeichnis

Adams M. D. *et al.* (2000). The Genome Sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195

Anderson S. (1981). Shotgun DNA sequencing using cloned DNase I-generated fragments. *Nucleic Acids Res.* **9**, 3015-3027.

Birnboim H.C. & Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* **7**, 1513-1523.

Bork P. and Copley R. (2001). The darft sequences: Filling in the gaps. *Nature* **409**, 818-820.

Bouck J., Miller W., Gorrell J. H., Muzny D. and Gibbs R. A. (1998). Analysis of the quality and utility of random shotgun sequencing at low redundancies. *Genome Res.* **8**, 1074-1084.

Bullock W.O., Fernandez J.M. & Short J.M. (1987). XL1-blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Bio Techniques* **5**, 376-379.

Cohen S.N. and Miller C.A. (1969). Multiple molecular species of circular R-factor DNA isolated from *E. coli*. *Nature* **224**, 1273-1277.

Collins F. and Galas D. (1993). A new five-year plan for the U.S. Human Genome Project. *Science* **262**, 43-46.

Collins F. S., Patrinos A., Jordan E., Chakravarti A., Gesteland R., Walters L. and the members of the DOE and NIH planning groups (1998). New Goals for the U.S. Human Genome Project: 1998-2003. *Science* **282**, 682-689.

Deininger P. L. (1983). Random subcloning of sonicated DNA; application to shotgun DNA sequence analysis. *Anal. Biochem.* **129**, 216-223.

Bird A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends Genet.* **3**,12, 342-347.

Burke D. T., Carle G. F. and Olson M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.

Dunham I., Shimizu N., Roe B. A., Chisoe S. et al. (1999). The DNA sequence of human chromosome 22. *Nature* **402**, 489-495.

Durland R.H. and Eastman E.M. (1998). Manufacturing and quality control of plasmid-based gene expression systems. *Adv. Drug. Del. Rev.* **30**, 33-48.

Fleischmann R. D. et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496-512.

Gardiner-Garden M. and Frommer M. (1987). CpG Islands in Vertebrate Genoms. *J. Mol. Biol.* **196**, 261-282.

Gardner R. C., Howarth A. J., Hahn P., Brown-Luedi M., Shepherd R. J. and Messing J. (1981). The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. *Nucl. Acids. Res.* **9**, 2871-2888.

Guttman A., Cohen A. S., Heiger D. N. and Karger B. L. (1990). Analytical and Micropreparative Ultrahigh Resolution of Oligonucleotides by Polyacrylamide Gel High-Performance Capillary Electrophoresis. *Anal. Chem.* **62**, 137-141.

Hengen P.N. (1997), Shearing DNA for genomic library construction. *Trends-Biochem.Sci.*; **22**, 7, 273-274.

Hodall E., Boye K. and Vuust J. (1999). Simple Preparation Method of PCR Fragments for Automated DNA Sequencing. *J. Cell. Biochem.* **73**, 4, 433-436.

Hoheisel, J.D., Drmanac, R., Larin, Z., Lennon, G., Manaco, A.P., Nizetic, D., Ross, M., Zehetner, G., Lehrach, H. (1991). Use of high coverage libraries for an integrated analysis of genomic DNA. *Adv. Mol. Gen.* **4**, 125-132.

Hoheisel J.D., Ross M.T., Zehetner G., Lehrach H. (1991). Relational genome analysis using reference libraries and hybridisation fingerprinting. *J. Biotech.* **35**, 121-134.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Jeffreys A. J., Wilson D. and Thein S. L. (1985). Hypervariable ' minisatellite' regions in human DNA. *Nature* **314**, 67-73.

Ju J., Ruan C., Fuller C. W., Glazer A. N. and Mathies R. A. (1995). Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci.* **92**, 4347-4351.

Kilger C. and Schmid K. (1994). Rapid characterization of bacterial clones by microwave treatment and PCR. *Trends Genet.* **10**, 5, 149.

Knippers R. (1985). Molekulare Genetik. 4. Auflage, Georg Thieme Verlag Stuttgart- New York.

Koop B.F., Rowan L., Chen W.Q., Deshpande P., Lee H., Hood L. (1993). Sequence length and error analysis of Sequenase and automated Taq cycle sequencing methods.. *Bio-Techniques* **14**, 442-447.

Krishnan B. R., Blakesley R. W. and Berg D. E. (1991). Linear amplification DNA sequencing directly from single phage plaques and bacterial colonies. *Nucleic Acids Res.* **19**, 5, 1153.

Lee L. G., Spurgeon S. L., Heiner C. R., Benson S. C., Rosenblum B. B., Menchen S. M., Graham R. J., Constantinescu A., Upadhyya K. G. and Cassel J. M. (1997). New energy transfer dyes for DNA sequencing. *Nucl. Acad. Res.* **25**, No. 14, 2816-2822.

Levy M.S., Collins I.J., Yim S.S., Ward J.M., Titchener-Hooker N., Ayazi-Shamlou P. and Dunnill P. (1999). Effect of shear on plasmid DNA in solution. *Bioprocess Eng.* **20**, 7-13

Levy M.S., Collins I.J., Tsai J.T., Ayazi-Shamlou P., Ward J.M. and Dunnill P. (2000). Removal of contaminant nucleic acids by nitrocellulose filtration during pharmaceutical-grade plasmid DNA processing. *J. Biotechnol.* **76**, 2-3, 197-205.

Liu Y., Bai J., Liang X. and Lubman D.M. (1995). Use of nitrocellulose film substrate in matrix-assisted laser desorption/ionization mass spectrometry for DNA mapping and screening. *Anal. Chem.* **67**, 3482-3490.

Luckey J. A., Drossman H., Kostichka A. J., Mead D. A., D'Cunha J., Norris T. B. and Smith L. M. (1990). High speed DNA sequencing by capillary electrophoresis. *Nucl. Acad. Res.* **18**, No. 14,

Lui M., Tempst P. and Erdjument-Bromage H. (1996). Methodological analysis of protein-nitrocellulose interactions to design a refined digestion protocol. *Anal. Biochem.* **241**, 156-166.

Maier E., Crollius H.R., Lehrach H. (1994). Hybridisation techniques on gridded high density DNA and *in situ* colony filters based on fluorescence detection. *Nucl. Acad. Res.* **22**, 3423-3424.

Marshall E. (1998a). A second private genome project. *Science* **281**, 1121.

Marshall E. (1998b). NIH to produce a "working draft" of the genome by 2001. *Science* **281**, 1774-1775.

Metzker M. L., Lu J. and Gibbs R. A. (1996). Electrophoretically Uniform Fluorescent Dyes for Automated DNA Sequencing. *Science* **271**, 1420-1422.

Morton N. E. (1991). Parameters of the human genome. *Proc Natl Acad Sci U S A* **88**, 7474-7476.

Moslemi A.-R., Lindberg C. and Oldfors A. (1997). Analysis of Multiple Mitochondrial DNA Deletions in Inclusion Body Myositis. *Hum.-Mutat.* **10**, 5, 381-386.

Murray V. (1989). Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucl. Acad. Res.* **17**, No. **21**, 8889.

Myers E. W. et al. (2000). A Whole-Genome Assembly of Drosophila. *Science* **287**, 2196-2204.

National Research Council. *Mapping and Sequencing the Human Genome* (National Academy Press, Washington DC, 1988).

Oefner P. J., Hunicke-Smith S.P., Chiang L., Dietrich F., Mulligan J., Davis R.; (1996), Efficient random subcloning of DNA sheared in a recirculating point-sink flow system. *Nucl. Acad. Res.* **24**, 20, 3879-3886.

Ohno K. et al. (1991). Direct DNA sequencing from colony: analysis of multiple deletions of mitochondrial genome. *Biochim. Biophys. Acta N* **1090**, 1, 9-16.

Olson M., Hood L., Cantor C. and Botstein, D. (1989). A common language for physical mapping of the human genome . *Science* **245**, 1434-1435.

Palca J. (1986). Human Genome-Department of Energy on the map. *Nature* **321**, 371.

Pennisi E. (1999). Academic sequencers challenge Celera in a sprint to the finish. *Science* **283**, 1822-1823.

Popovic and Wintzerith (1983). Hydrophobic interaction of ribonucleic acids on a nitrocellulose column. *J. Chromatogr.* **268**, 285-290.

Prazeres D., Schulep T., and Cooney C. (1998). Preparative purification of supercoiled plasmid DNA using anion-exchange chromatography. *J. Chromotogr. A* **806**, 31-45.

Prober J. A. et al. (1987). A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. *Science* **238**, 336-341.

Reeve M. A. and Fuller C. W. (1995). A novel thermostable polymerase for DNA sequencing. *Nature* **376**, 796-797.

Reinikainen P., Korpela K., Nissinen V., Olkku J., Söderlund H. and Markkanen P.; (1988). *Escherichia coli* Plasmid Production in Fermenter. *Biotechnol. And Bioeng.* **33**, 386-393.

Rosenblum B. B., Lee L. G., Spurgeon S. L., Khan S. H., Menchen S. M., Heiner C. R. and Chen S. M. (1997). New dye-labeled terminators for improved DNA sequencing patterns. *Nucl. Acad. Res.* **25**, No. **22**, 4500-4504.

Rubin G. M. et al. (2000). Comparative Genomics of the Eukaryotes. *Science* **287**, 2204-2215.

Rush M.G. and Warner R. (1970). Alkali denaturation of covalently closed circular duplex deoxyribonucleic acid. *J. Biol. Chem.* **245**, 2704-2708.

Saiki R. K., Scharf S., Faloona F., Mullis K. B., Horn G. T., Erlich H. A. & Arnheim, N. (1985). Enzymatic amplifikation of β -globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). Molecular Cloning: A laboratory Manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour.

Sanger F., Nicklen S. & Coulson A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**, 5463-5467.

Sender A.J. (2001). Roll over, Plasmid Prep. *Genome Technology* Juli 2001, 44-48.

Shizuya H. *et al.* (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794-8797.

Sinsheimer R. L. (1989). The Santa Cruz Workshop-May 1985. *Genomics* **5**, 954-956.

Smith L. M. *et al.* (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* **321**, 674-679.

Swerdlow H., Wu S. L., Harke H., Dovichi N. J. (1990). Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *J. Chromotogr.* **516**, 61-67.

Tabor S. and Richardson C. C. (1989). Selective Inactivation of the Exonuclease Activity of Bacteriophage T7 DNA Polymerase by *in Vitro* Mutagenesis. *J. Biol. Chem.* **264**, No. **11**, 6447-6458.

Tabor S. and Richardson C. C. (1990). DNA Sequence Analysis with a Modified Bacteriophage T7 DNA Polymerase. *J. Biol. Chem.* **265**, No. **14**, 8322-8328.

The chromosome 21 mapping and sequencing consortium (2000). The DNA sequence of human chromosome 21. *Nature* **405**, 311-319.

Thorstenson Y.R., Hunicke-Smith S.P., Oefner P.J., Davis R.W. (1998). An automated hydrodynamic process for controlled, unbiased DNA shearing. *Genome Res.* **8**, 848-855.

Venter J. C., Adams M. D., Sutton G. G., Kerlavage A. R., Smith H. O. and Hunkapiller M. (1998). Shotgun Sequencing of the Human Genome. *Science* **280**, 1540-1542.

Venter J. C. *et al.* (2001). The Sequence of the Human Genome. *Science* **291**, 1305-1351.

Vieira J & Messing J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with sequencing universal primer. *Gene* **19**, 259-268.

Wahl F. M. (1998). Digitale Bildverarbeitung. Nachrichtentechnik 13, Springer Verlag, 107ff.

Waterson R. H., Lander E. S. and Sulston J. E. (2001). On the sequencing of the human genome. “www.pnas.org/cgi/doi/10.1073/pnas.042692499”.

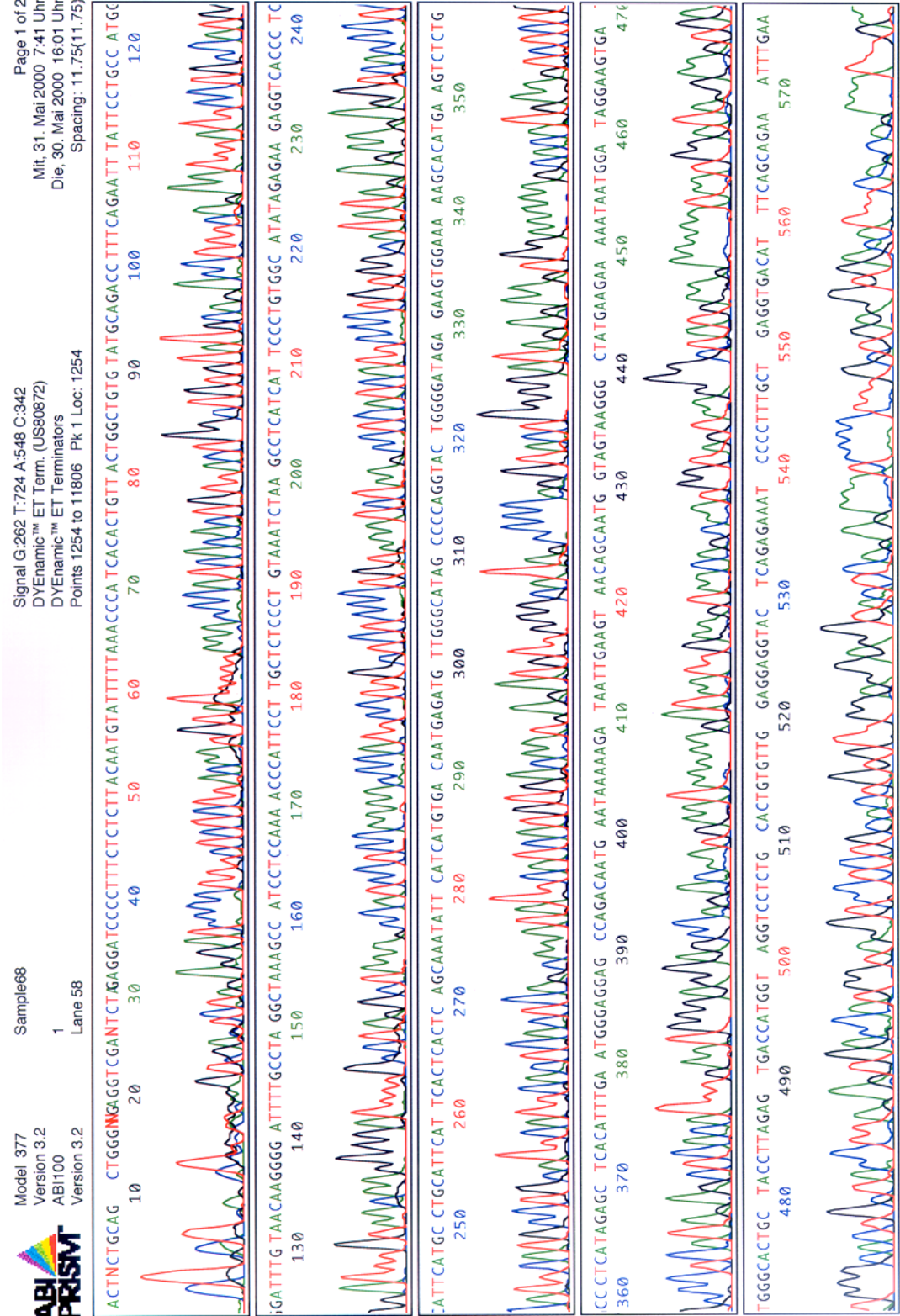
Wohlhieter J.A., Falkow S. and Citarella R.V. (1966). Purification of episomal DNA with cellulose nitrate membrane filters. *Biochim. Biophys. Acta* **129**, 475-481.

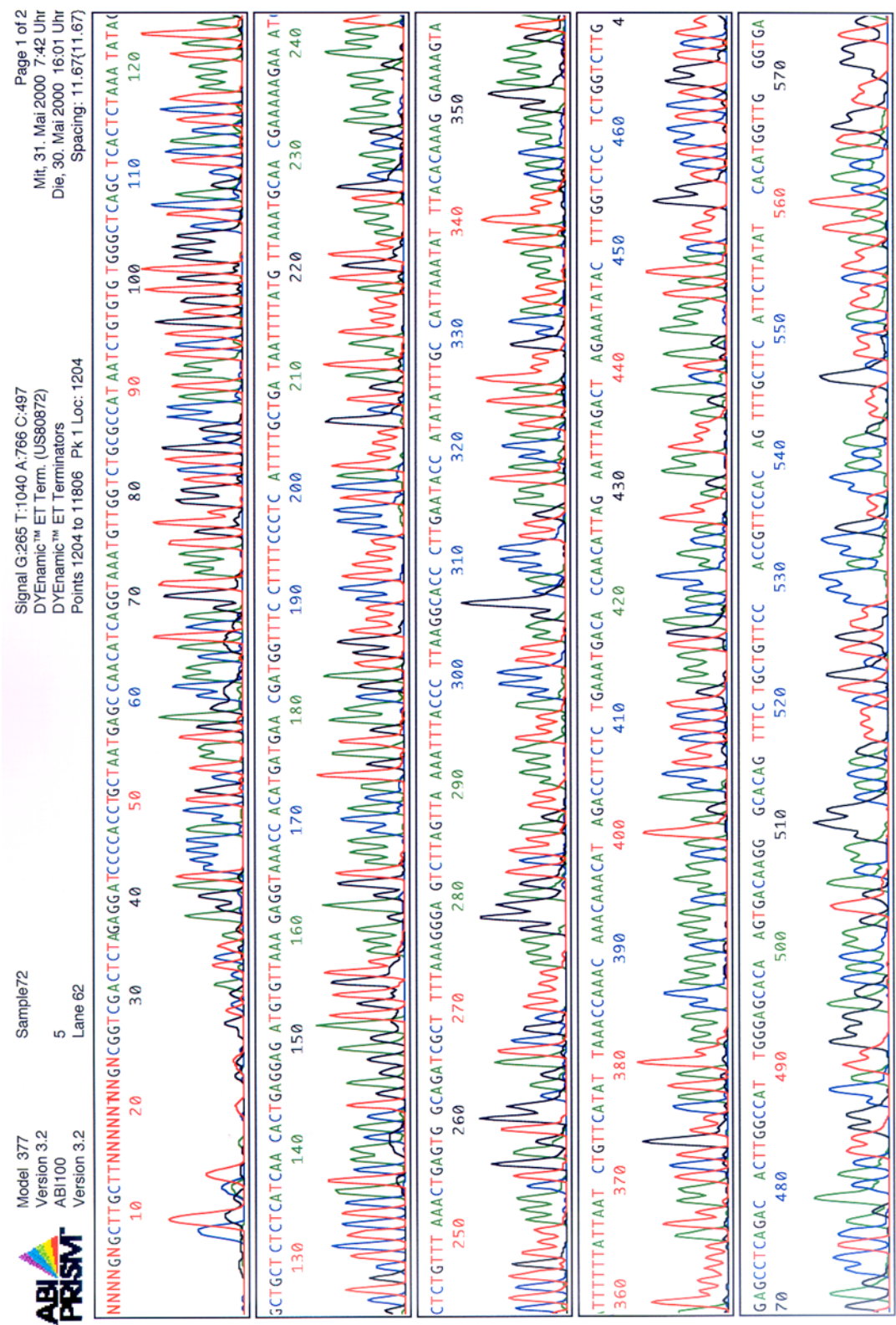
Wooster R. et al. (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378**, 789-792.

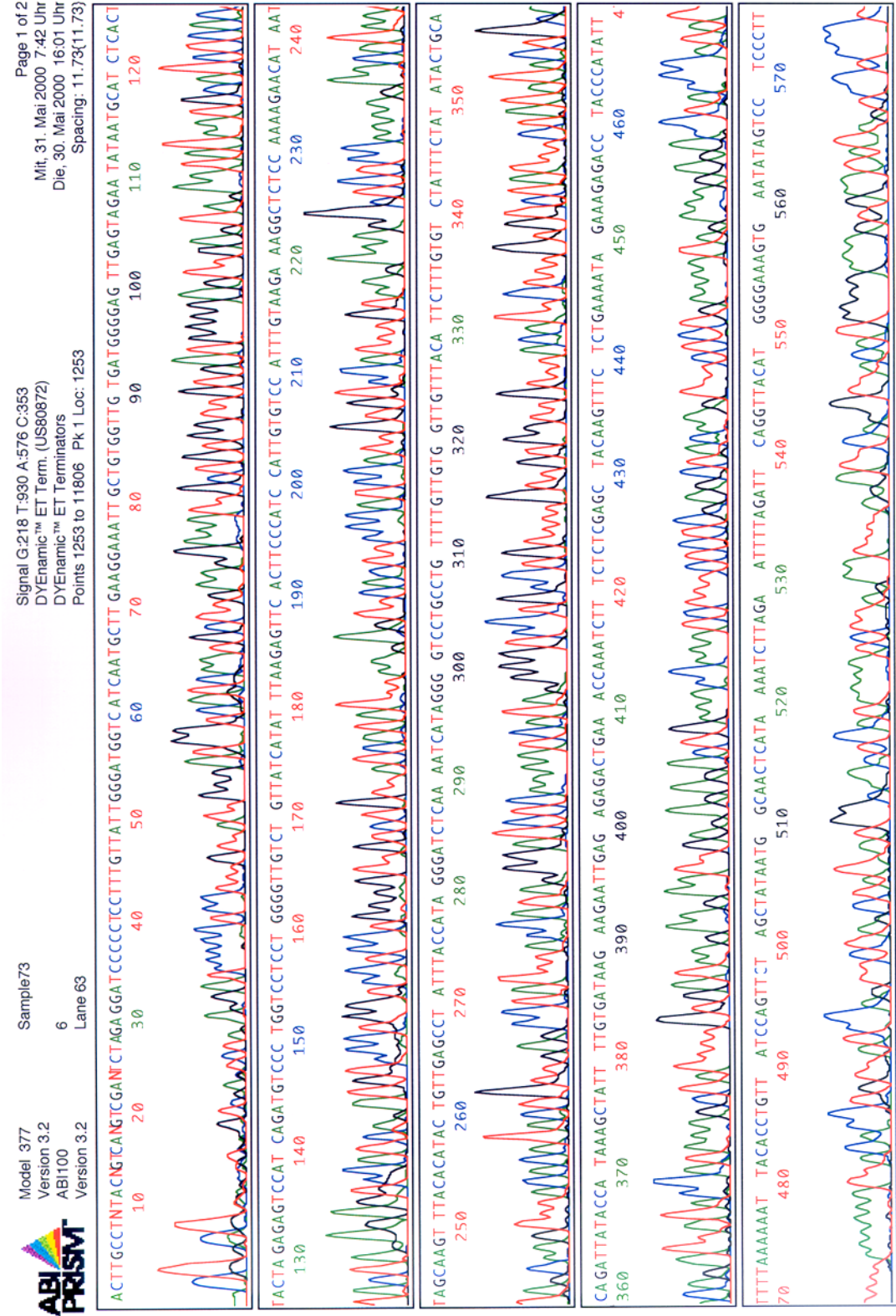
Ymer S. (1991). Heat inactivation of DNA ligase prior to electroporation increases transformation efficiency. *Nucleic Acids Res.* **19**, 6960.

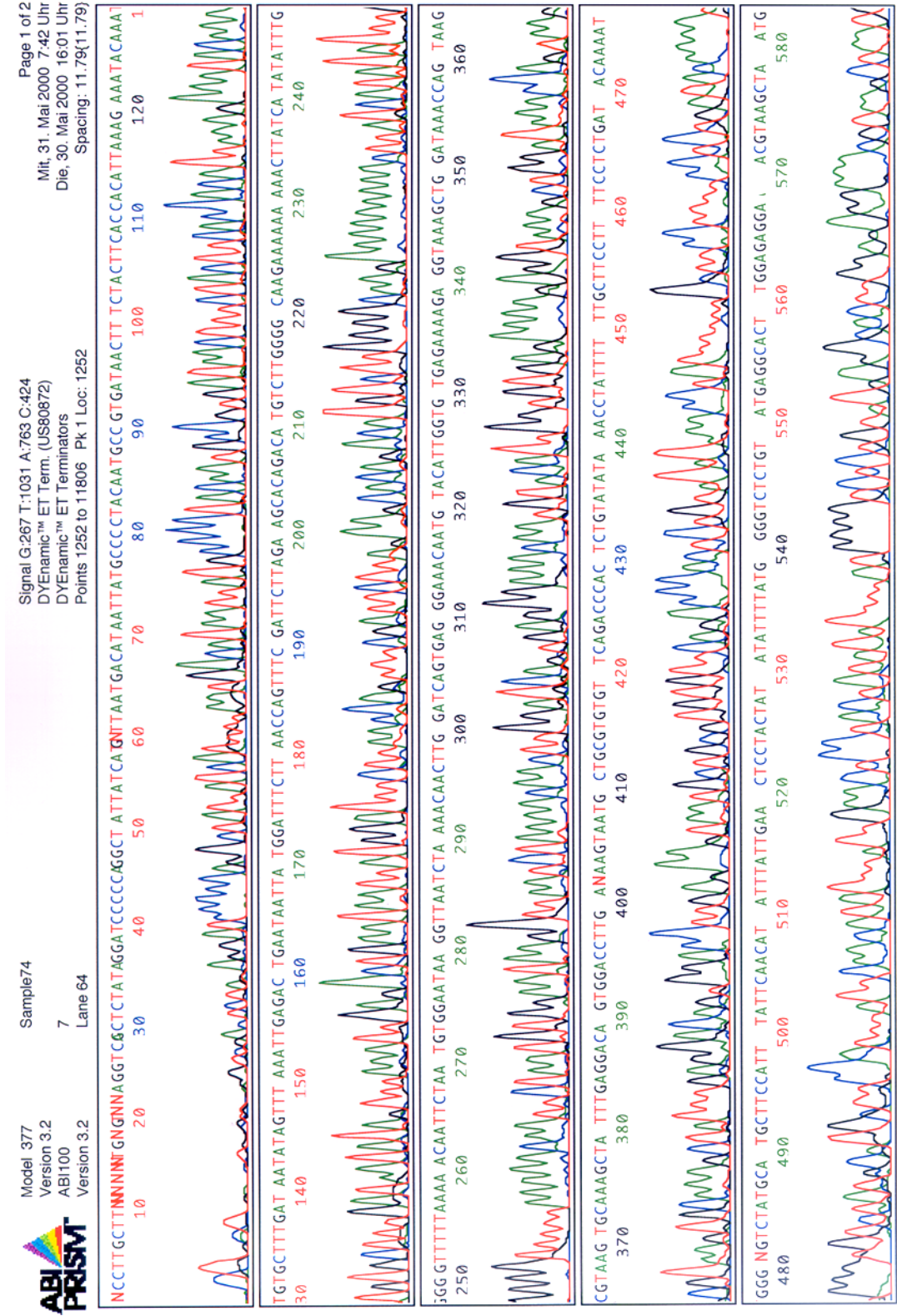
7 Anhang

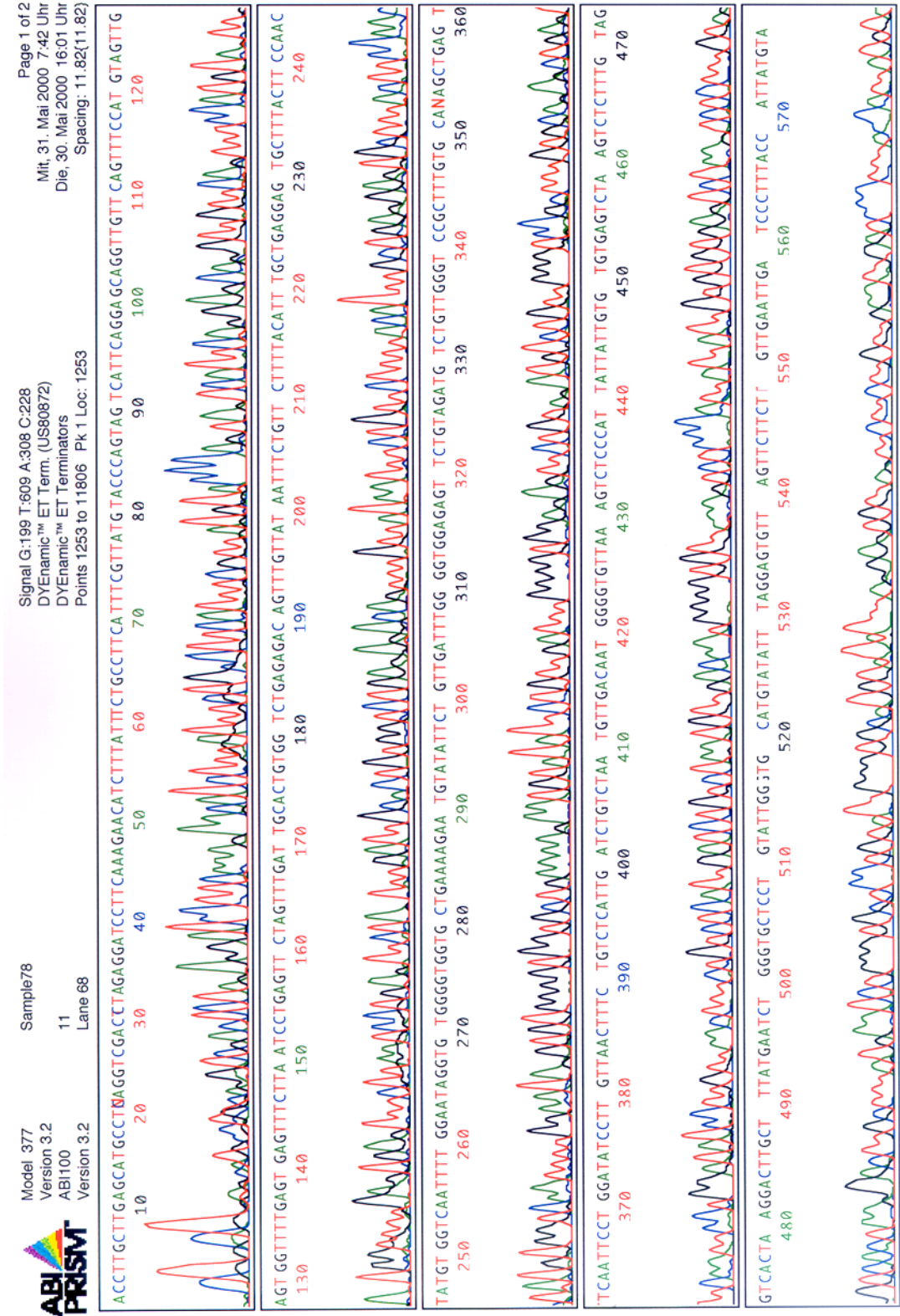
7.1 Anhang A: Elektropherogramme

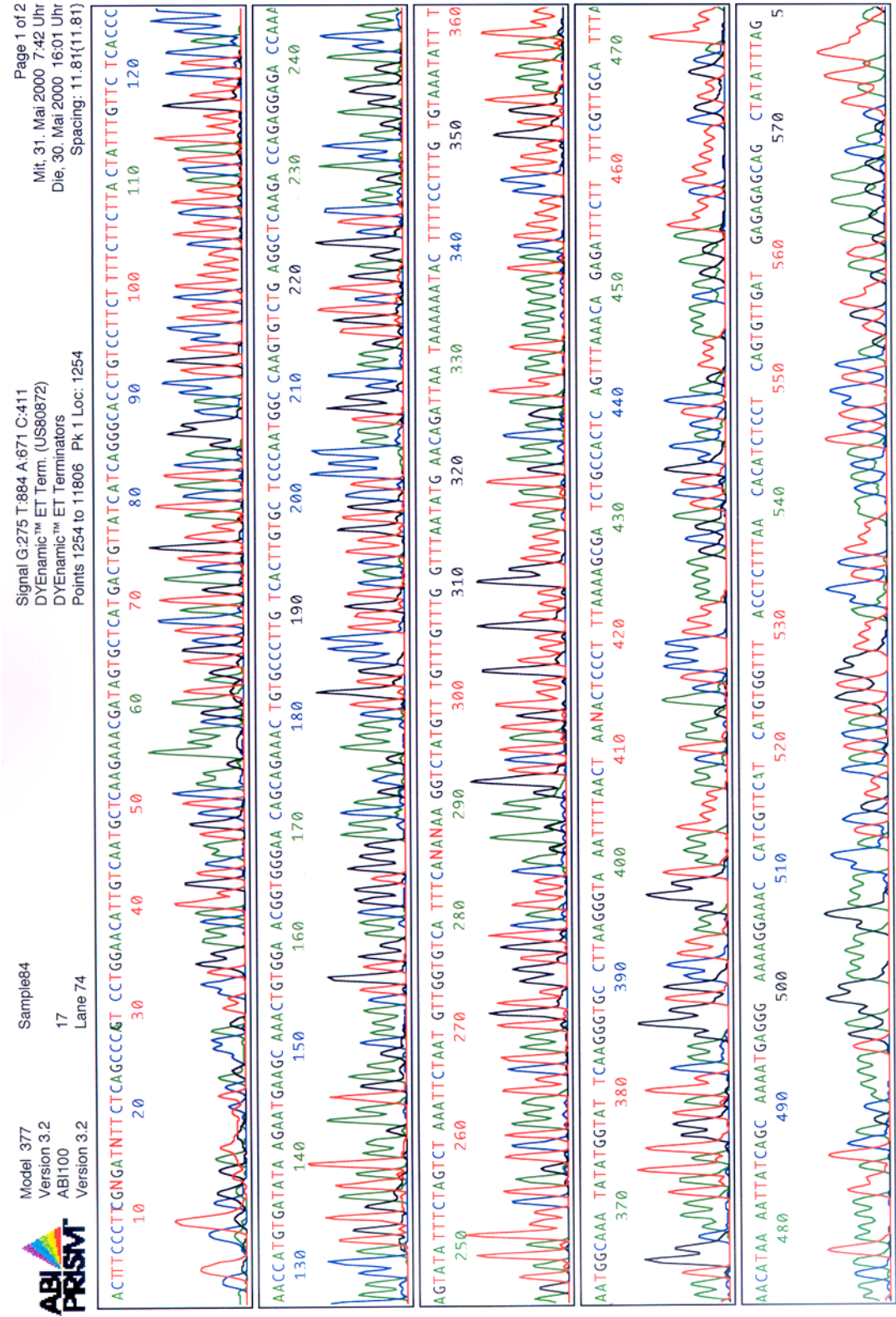












1020

1010

1000

990

980

970

960

950

940

930

920

910

900

890

880

870

860

850

840

830

820

810

800

790

780

770

760

750

740

730

720

710

700

690

680

670

660

650

640

630

620

610

600

590

580

570

560

550

540

530

520

510

500

490

480

470

460

450

440

430

420

410

400

390

380

370

360

350

340

330

320

310

300

290

280

270

260

250

240

230

220

210

200

190

180

170

160

150

140

130

120

110

100

90

80

70

60

50

40

30

20

10

AACATAAATATATCAGCAAAATGAGGGAAAAAGGAAACCATCGTTTCATCATGTGGTTTACCTCTTTAAACACATCTCCTCAGTGTGATGAGAGAGCAGCTATATTAG

1020

1010

1000

990

980

970

960

950

940

930

920

910

900

890

880

870

860

850

840

830

820

810

800

790

780

770

760

750

740

730

720

710

700

690

680

670

660

650

640

630

620

610

600

590

580

570

560

550

540

530

520

510

500

490

480

470

460

450

440

430

420

410

400

390

380

370

360

350

340

330

320

310

300

290

280

270

260

250

240

230

220

210

200

190

180

170

160

150

140

130

120

110

100

90

80

70

60

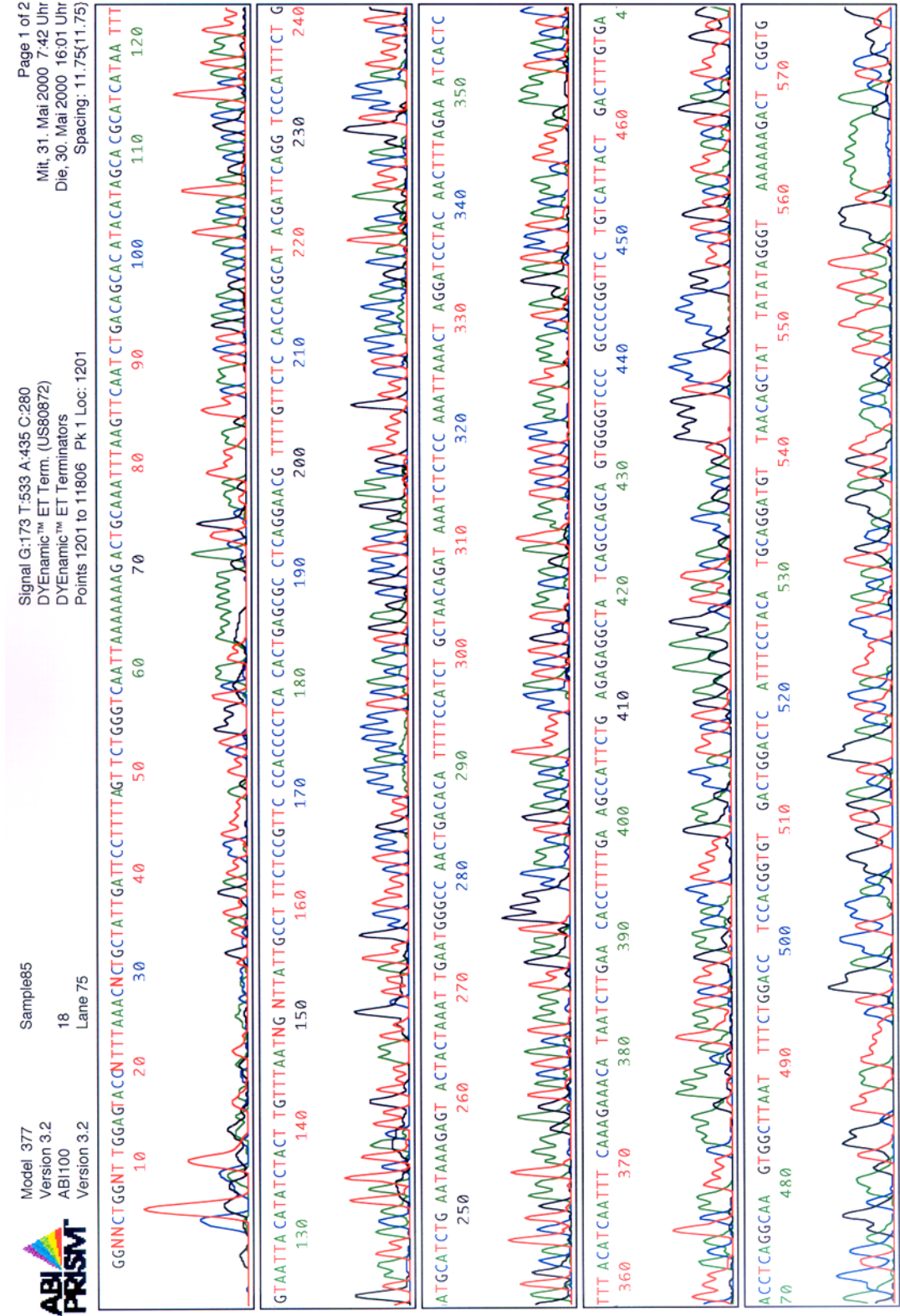
50

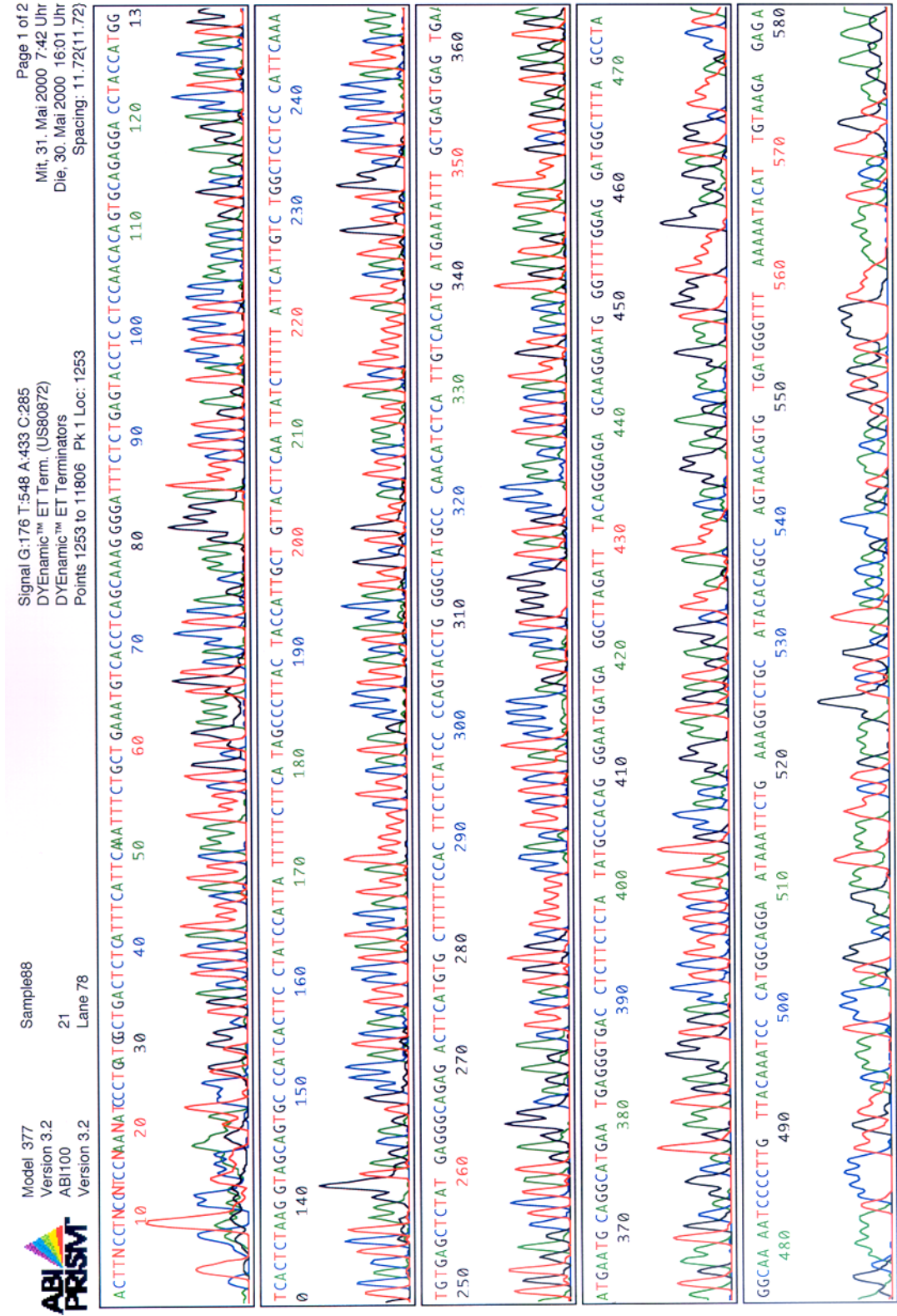
40

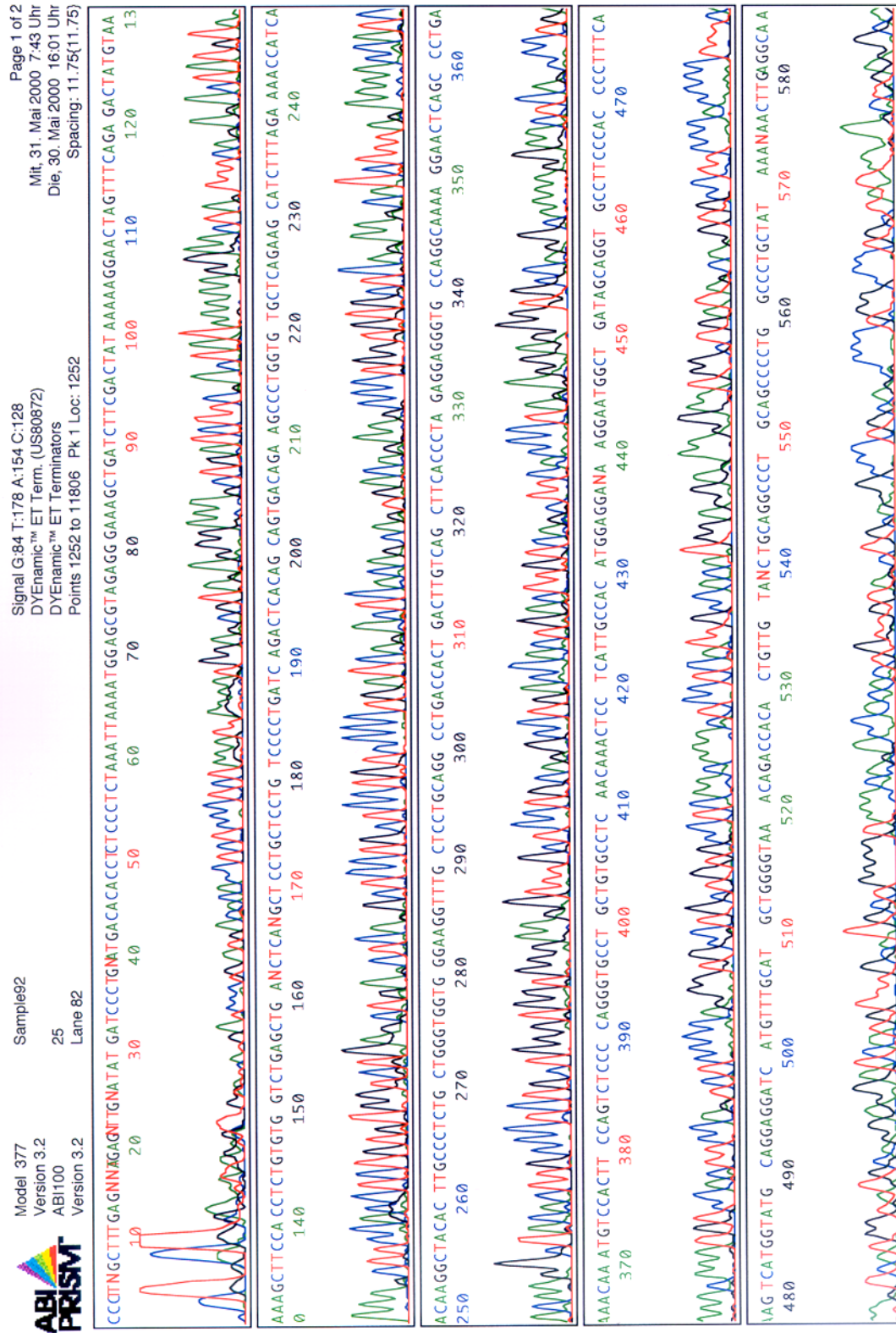
30

20

10







7.2 Anhang B: Verwendete Sequenzierprimer

Es handelt sich hierbei um eine Auflistung der Primer die zum Schließen der Lücken zwischen den einzelnen Contigs und zur Abdeckung einzelsträngiger Bereiche um die nach HUGO-Konvention erforderliche Abgabequalität zu erreichen.

7.2.1 Primer für den Klon 98L15:

h25_1c1.q3	5'-ATg Tgg Agg gTA ggg Agg-3'	18mer
h25_1g1.p3	5'-gTT gAA CAA AgA TgA gAg AgA gAA g-3'	25mer
h25_1g11.q3	5'-CAT AAC CCA gCg gTT AAA gTC-3'	21mer
h25_2m3.R2	5'-ACT TTg AgA TTA Aag gAA ATT Cag-3'	24mer
h25_2m3.p3	5'-gTT CTT CCC Tag Tgg TAA TgA gAT g-3'	25mer
h25_4d15.q3	5'-CCT ATT Agg ggC TCA TgT TTT TC-3'	23mer
h25_4d21.p3	5'-ggT ggT TgT gTT TCT gTg gg-3'	20mer
h25_4f5.q3	5'-TCT CTg gCT TTC AAC TgA Agg-3'	21mer
h25_4h3.R2	5'-TgA gTg ATA gAA gTT CAT TAT AAT AT-3'	26mer
h25_4h15.q3	5'-CAA CCA CAA TCC TAC TTT CCg TC-3'	23mer
h25_4j11.q3	5'-Tgg Cag Agg CAA CTC AgA TAA C-3'	22mer
h25_4p1.q2	5'-CCA ACA Cag gAg CAC CCA g-3'	19mer
h25_5d3.q3	5'-AAA gTT Cag TCA CAC AgC TCC-3'	21mer
h25_5n24.p2	5'-CCA TgT Tag TgT gCT gCA CCC-3'	21mer
h25_6a19.q3	5'-AAA CTA ATC TgA ggT CTg ggg-3'	21mer
h25_6a19.q4	5'-ggA AAA TAT Tgg ggg Atg ACA Aag-3'	24mer
h25_6m9.q3	5'-CAA ggA TgA CgA TTg ACA Aag g-3'	22mer
h25_6m9.q4	5'-gAA Cag AgA gAg Aag TCg gAA Aag-3'	24mer
h25_7m12.p3	5'-CCA CCA TTT ggA ggA TAA TAT TTg C-3'	25mer
h25_10c13.q3	5'-CTC gTT ATC TgA gTT gCC TCT g-3'	22mer
h25_10e9.p3	5'-gCT AAG TTC AAA gAA CTg TCC C-3'	22mer
h25_10e21.R2	5'-gAT gAA TgT CTT TgC TCA TAT ATT T-3'	25mer
h25_10e21.R3	5'-CAT CTT TCT CAT ATT ggC AgT Agg-3'	24mer
h25_10k13.q2	5'-ggA TAT TTT gAg Aag Tgg gTA Tgg-3'	24mer
h25_10m11.p3	5'-gAA TgC TgA AAC CTC CTg AAC-3'	21mer
h25_11c14.q3	5'-Cag TCT ACA AgT TCA AAA TgC CC-3'	23mer
h25_12f15.S2	5'-TTT ggg CTA TTA TTT gTA ATC CTC-3'	24mer
h25_12f15.S3	5'-AAC AAC CCC CCC AAC TgA AAA C-3'	22mer
h25_12n15.q2	5'-CAA TAA CTA Agg AgA ggC AgA AAC-3'	24mer

7.2.2 Primer für den Klon 49E3:

ABfor.p	5'-CCA TCA TCA CAT CAC CAT CAA CAC TAT-3'	27mer
AB0.p	5'-CgT CAg AgA gAC CTT CCT gTA TT-3'	23mer
A-B1.p	5'-CTg CCT CAg gCT CCC AAA-3'	18mer
A-B2.p	5'-ggA TCC TgC TCT gTg TgT CTT T-3'	22mer
A-B3.p	5'-CCA TCA TCA TCA ACA TCA CCA T-3'	22mer
A-B4.p	5'-CCA TCA CCA CCA TCA CCA TAA T-3'	22mer
AB0.q	5'-ggT CCC ACT TTC TCA CTT gTA AAA-3'	24mer
A-B1.q	5'-CCT ACA TTA TgA AgT TgT TAC gAg AA-3'	26mer
A-B2.q	5'-gTg AT AgT gTT gAT ggT gAT gTg AT-3'	25mer
AB2a.q	5'-CAT Agg CTC Tgg AAT gAg ACT gCT TA-3'	26mer
A-B3.q	5'-ggA TgA Tgg CTg ACA gCA AgT A-3'	22mer
A-B4.q	5'-gTg gTg ggC ACC TgC AAT-3'	18mer
B-C1.p	5'-ggA AgA Agg CAT CTT ATg TgT TCT T-3'	25mer
B-C2.p	5'-CAg gCC ATT CCT TCC ATC AT-3'	20mer
B-C3.p	5'-ggT TCA CAA TgC CTA TCA ATg TT-3'	23mer
B-C4.p	5'-CTC CCT AAg gCT TTT gCT CAA-3'	21mer
B-C5.p	5'-CCC AgC CTC Tgg TAA CCA T-3'	19mer
B-C6.p	5'-CAT ggA AgT gCA gAT ATC TCT TCA A-3'	25mer
B-C7.p	5'-CTT ggA gAT TCA CCC CTT ATC ATA-3'	24mer
B-C1.q	5'-ggg TAC ACA ggA AgC CCA AA-3'	20mer
B-C2.q	5'-ggT gTT ggA AAA ACT gTA TAg TCA CA-3'	26mer
B-C3.q	5'-CAA gTg gCC AAC Agg TAT ATg AA-3'	23mer
B-C4.q	5'-CCC ATC ATT ggC AAC AgC AT-3'	20mer
B-C5.q	5'-gAT gTg gTg ATA ggT TgA gCA AA-3'	23mer
B-C6.q	5'-gCC Tgg gAg AAg gAA CTA TTA A-3'	22mer
B-C7.q	5'-ggC TTC TTg gAA AgC TTg AAA-3'	21mer
B-C8.q	5'-CCC TAC ACA gCC AAA TgA gAT A-3'	22mer
BCrev.q	5'-gTg CAg CAT ggT GAC CAC AgT TA-3'	23mer
C-D1.p	5'-CTg gAC CAT AAg gCA TCA CAT T-3'	22mer
C-D2.p	5'-CAg ACC TgT TTg TCT TTT CCC TT-3'	23mer
C-D3.p	5'-CTg CAg TgT Agg CTg AAT gAT gTA-3'	24mer
C-D4.p	5'-gTA gCC TgT gTA TCA ACT CTg gAT T-3'	25mer
C-D5.p	5'-CAT gCA TgT Tgg AgA AAA CTg AA-3'	23mer
CDfor.p	5'-CAT CCA AAg CAT gCT AgT TGgg AgA AA-3'	26mer
C-D1.q	5'-gCT ACC ATT CTC ACT CTg AgA AgT AT-3'	26mer
C-D2.q	5'-ggA Agg CTC TgA TCA CAg TTg A-3'	22mer
C-D3.q	5'-CgT gTA CCT AAC ATT gAT gAT TTC TT-3'	26mer

Anhang

C-D4.q	5'-CCC ATT gTg gAA gAT TTg gAA-3'	21mer
C-D5.q	5'-gAC gTC AAg gAg gAg gAA CAT T-3'	22mer
C-D6.q	5'-ggg TgA AgT CAT TA gAT ATg CAA-3'	24mer
CDrev.q	5'-CAA TgA ACA ggC TgA CCA CgT CAT A-3'	25mer
D-E1.p	5'-gAA ggT Agg AAC AAA gAg ATT TCA T-3'	25mer
D-E2.p	5'-gAA CCC Agg TTT CTT ggC TTA-3'	21mer
D-E1.q	5'-gAT gCT TgA ggA AgA ATA TAA Tgg AA-3'	26mer
D-E2.q	5'-gAT CTg gAA TAA ATg ACA AgT gAA A-3'	25mer
E-F1.p	5'-CTC Agg gAT CTg Tgg gAC TAT AA-3'	23mer
E-F2.p	5'-CTg CCA ACC TCA AAT CTT ACA T-3'	22mer
E-F3.p	5'-ggC CTA CAT ggA AgT gAg gTT T-3'	22mer
E-F4.p	5'-CTC AgC CCA CTA ATg TgA ATg AAA-3'	24mer
E-F5.p	5'-ggA AAg CCT AgA TTT CCT TgA AT-3'	23mer
E-F6.p	5'-CTC CTT ggT gCT CAT CAT gAA-3'	21mer
E-F7.p	5'-Cgg AgA TCC ACC Agg CTT T-3'	19mer
E-F8.p	5'-CCC TgT CTC ATC ACT gTA TTT gA-3'	23mer
E-F9.p	5'-ggT CAg TAA CAg ACA ACA ggA AAA-3'	24mer
E-F10.p	5'-gTC TA g CAA gAC AgA CAA AAT AAA A-3'	25mer
E-F11.p	5'-gCC AAA ACC AgA CAA AgA CAA T-3'	22mer
E-F12.p	5'-gCC ATg gCA gCA AgT AgA TT-3'	20mer
E-F13.p	5'-CCC gAA TTg AAA gAT gTg gTT AA-3'	23mer
E-F14.p	5'-ggA TTT CAg ATg TAA ATg TAA CAC ATA-3'	27mer
E-F15.p	5'-gCC TgT CAg AAT gCg TAA gAA T-3'	22mer
EF15a.p	5'-CCA gCA ATA ACA CTC CTg ggA ATT-3'	24mer
E-F16.p	5'-ggT ggC ACA TCC TgT ATC TAT gTA A-3'	25mer
E-F17.p	5'-ggA ggT TCA TCC CTg TA g CAT AT-3'	23mer
E-F18.p	5'-gTC ACA TgC Tgg ATg ATT CTC TTT-3'	24mer
E-F19.p	5'-ggC AgT gAC TgT gAC Tgg AA-3'	20mer
E-F20.p	5'-gCT CCC TAA TTC TCC AAT ACg TAT A-3'	25mer
EFfor1.p	5'-CCA gAT gCT CTT CAA CAg gTg ATT-3'	24mer
EFfor2.p	5'-ggT ggC ACA TCC TgT ATC TAT gTA AA-3'	26mer
E-F1.q	5'-CCC TTT ACA TTC CCC ACA TTT-3'	21mer
E-F2.q	5'-gTC TTC CTC TTC CTC TTC CTC TTT-3'	24mer
E-F3.q	5'-ggT ggC TTA AAA CAA CAC AAA TAT-3'	24mer
E-F4.q	5'-ggA TTC TTC TTC TAC CAg TTT CCA A-3'	25mer
E-F5.q	5'-ggC TAA Agg AAT CTg TgA gTC AT-3'	23mer
E-F6.q	5'-CTC CCT ggg TTC ATg gCT T-3'	19mer
E-F7.q	5'-CAA ggT TgA TCT CAg ACA TCA TCA A-3'	25mer

Anhang

E-F8.q	5'-ggC CTC AgA TAT TTC TTC ATT TCT A-3'	25mer
E-F9.q	5'-CCA TTT ggg TCT ggA gAg TTT-3'	21mer
E-F10.q	5'-CCC TCT ATT CCT AAC AAg CCA A-3'	22mer
E-F11.q	5'-CAg ATC CTg TgC ATC TTT TgT T-3'	22mer
E-F12.q	5'-gTA CCT ATT TgT gTT ggg TCT gTA TT-3'	26mer
E-F13.q	5'-ggT gAA ATg TgT ATT CgT gTC TTT-3'	24mer
E-F14.q	5'-CCA ATC ACC TgT TgA AgA gCA T-3'	22mer
E-F15.q	5'-CCT ATg ggA TAC TAC TCA gCA ACT AA-3'	26mer
EF15a.q	5'-ggA ggT TCA TCC ATg TAg CAT gTA T-3'	25mer
E-F16.q	5'-ggC TCT gTA TCT ACC ACC ACA AT-3'	23mer
E-F17.q	5'-gTg AAg AAT CTT ATg gAg gAA gTT T-3'	25mer
E-F18.q	5'-ggT TgC ATT CCT TCA TCC TAA A-3'	22mer
E-F19.q	5'-gAg gTA gCA ATg Tgg gAA CAA A-3'	22mer
E-F20.q	5'-gAA CTT gAg TAA AAC CTC TgA ATT CT-3'	26mer
EFrev1.q	5'-CAg ggA TgA ACC TCC Agg gAA TT-3'	23mer
EFrev2.q	5'-ggC TCT gTA TCT ACC ACC ACA ATg AA-3'	26mer
FG1.p	5'-CCC TgT TAg AgA TTT TCA gCT AAg AT-3'	26mer
FG2.p	5'-ggA TgA CAT gAC ACA TCT TgA AA-3'	23mer
FG3.p	5'-ggA AgT gAA TgA ATA gCA ggT AAC TT-3'	26mer
FG4.p	5'-CAT gAC TTT CTg ggA ATT gTC AA-3'	23mer
FG5.p	5'-gCT CAT ggg TAg gAA gAA TCA AT-3'	23mer
FG6.p	5'-CTg gAT CCC TTC CTT ACA CCT TA-3'	23mer
FG7.p	5'-gAg CAg ACA CTT CTC AAA AGA AgA-3'	24mer
FG1.q	5'-CCA TgT gTT CTC ACT gTT CAA CAA-3'	24mer
FG2.q	5'-gTC CCA CCA ACA gTg TAA AAg TAT T-3'	25mer
FG3.q	5'-gCT gTg CAg AAg CTC TTT AgT TT-3'	23mer
FG4.q	5'-gCT CTg TTC TgT TCC ATT ggT CTA-3'	24mer
FG5.q	5'-gAg gTC CTT CAC ATC CCT TgT AA-3'	23mer
FG5a.q	5'-gAg gTC CTT CAC ATC CCT TgT AAg TT-3'	26mer
FG6.q	5'-gAg Agg ACC AgA Tgg TAT TCA TTA T-3'	25mer
FG6a.q	5'-CTC TCT gTT TgT CTg TTA TTg gTg CTA-3'	27mer
FG7.q	5'-ggC CAA ggT TCT CCA gAA AA-3'	20mer
FGrev1.q	5'-CCT TCA CAT CCC TTg TAA gTT ggA TT-3'	26mer
FGrev2.q	5'-CCC AgA AAg TCA TgA gCC TAg CAT T-3'	25mer
F19.p	5'-gCT CCT TAA TAT gTT ACA TgC CAA A-3'	25mer
19.q	5'-ggA Tgg AgT ATg CTA AgC TCT gAT AA-3'	26mer
1374_1.q	5'-CAT ggT TCA TAC CTT CTA AAT ggA TA-3'	26mer
1374_2.q	5'-gCC AAA ggg gTT CgA ggA ATA-3'	21mer

Anhang

F1374.q	5'-gTg ggC AgA TTA gAT ggC TTT ATA-3'	24mer
F1498.p	5'-gTA CAT AAC gAA ATg AAg gCA gAA AT-3'	26mer
1498.q	5'-gCC ACA ACT TCA gAg CCT gTT A-3'	22mer
1516.q	5'-gAT gCC TCC AAC TTT gTT CTT TT-3'	23mer
F1516.q	5'-CTg AgA gCT CTg TTC TgT TCC ATT-3'	24mer
F1533.p	5'-ggA ACT gTA AAC TAg TTC AAC CAT T-3'	25mer
1533.q	5'-gTg CAC AAT gTg CAg gTT AgT TA-3'	23mer
F1533.q	5'-gAC TgC AAg gAA TTT TgC TCT ATT-3'	24mer
F1638.p	5'-gTg CAC TAC gTg CAg gTT TgT TA-3'	23mer
1638.p	5'-CCg TTA ACg CCT CAT TTA CAT TA-3'	23mer
1655_2.q	5'-ggg TAT CTA gCA AgA AAA gAA gCA A-3'	25mer
1655_1.q	5'-ggA Agg AAg AAC AAT gAA CTA gTT AA-3'	26mer
F1655.q	5'-gAg AgA AAC TgT CAT gAg CAg AAA-3'	24mer
1046_2.q	5'-ggT Tgg TTC CAA gTC AAA gCT ATT-3'	24mer
1046_1.q	5'-CCA CCA AAC ATg gAg gTT CTA TA-3'	23mer
F1046.q	5'-CAA CCC TAg TgC TTA AAg CTC ATT-3'	24mer
F737.p	5'-CAT CAC CAT gTT AAg TAA gCA CTT T-3'	25mer
738.q	5'-CCA gAg AAT TgA TTT CAg CAA CTA T-3'	25mer
F738.q	5'-CTC CTC AgC CAC CTA Tgg TTT A-3'	22mer
F1043.p	5'-gAg gAA gCA gAA AAA gTA AAC AgA A-3'	25mer
2100_2.q	5'-CAg CCA AAT gAg ATA gAA Tgg AAA-3'	24mer
2100_1.q	5'-CAg TgA CTT ggC TTC TTg gAA A-3'	22mer
F2100.q	5'-gTg ATg CTT gTA gAg gCT TTg TAA-3'	24mer
2082_2.q	5'-gCA gAg TAg ATg TTg AAT gCA AAA-3'	24mer
2082_1.q	5'-gCA CTT CCA TgT TAA TTA CAg CAT T-3'	25mer
465_5.q	5'-ggA ACA gAA CAg AgA gCC TAg AAA-3'	24mer
465_4.q	5'-CAg AAA ACT gTA AAA CAT TgA TgA A-3'	25mer
465_3.q	5'-gCT ggA AAC CTT ATC CAA gTA ACA T-3'	25mer
465_2.q	5'-CCA gCT TCA CAA TgC TTA gTA ATT-3'	24mer
465_1.q	5'-CTA CTT CTC CTC TgA ggA TTC CAA T-3'	25mer
F465.q	5'-CCA TgC TgA ACT gTg AgT CAA TT-3'	23mer
F484.p	5'-CCA gCC AAg AAg AAA AgT AgA TTA-3'	24mer
495.q	5'-ggg CAA TCA gAg ATA AgA gAg AT-3'	23mer
F495.q	5'-CTg CTC TgT gAC TCT CCT gCA A-3'	22mer
F537.p	5'-ggg CTg CTA AAC TTg CTg AgA TA-3'	23mer
543.p	5'-CAg ggA ATC AgT gCA Tgg TTA A-3'	22mer
F544.q	5'-gTT CAg CAC CTg TgT CTA CCA TAA-3'	24mer
28n24.p2	5'-CTA CAT CCA Tgg ATT CAA CCA ATA-3'	24mer

Anhang

28n24.p3	5'-CAg CAT CAg gAA gTT TTT ATg ATT-3'	24mer
21i13.q3	5'-gAA CAg gCT gAC CAC gTC ATA A-3'	22mer
21i13.q2	5'-CAC TAA CTg TTg Tgg gAC ACA ATT-3'	24mer
21k1.q2	5'-gAA CTA TCA CAA CAT AAT gAT ggC TA-3'	26mer
21k1.p2	5'-CCA gAC TTT CAg Tgg AAA CCA T-3'	22mer
29m9.p3	5'-CCT gAA CCC CCT AAA CAT AgA TAA-3'	24mer
29m9.q2	5'-CAC TAT CAA TCC TCT TAC CTg gAA TT-3'	26mer
29m9.q3	5'-CCT ggT ACA TAT Tgg gCA CTA AA-3'	23mer
6i20.p2	5'-gCA CAg TTC AAg gCA CTT gAA AT-3'	23mer
6i20.p3	5'-ggT CAT gTg gAg TTT gTA AgC TAT T-3'	25mer
28j2.q2	5'-gAT ggg ACA ACT gAT TgC AgT TAA-3'	24mer
28j2.q3	5'-CCT ggA ACT AAA AgA ACg AAA AgA A-3'	25mer
28j2.q4	5'-gTA gAT gTT gCT Tgg TTg TTC CTA A-3'	25mer
6k4.p2	5'-ggA ggC CAC AgA gAA CAA AAT T-3'	22mer
6k4.p3	5'-CTA ggT gCT TAA Cgg ACA ACT TAA A-3'	25mer
23L11.p2	5'-gTC TgT gAA gCA AgA TCC ATA TTC AT-3'	26mer
23L11.p3	5'-gAg CCA TCA CAA TAA TgA TTC AAA-3'	24mer
23L11.p4	5'-CCA gTC AAT CTT AAT gAC AgA gAg AA-3'	26mer
2100_b.q	5'-gTC TAT TTC ATC CCT ACA CAg CCA AA-3'	26mer
2100_a.q	5'-CTT ggC TTC TTg gAA AgC TTg AAA T-3'	25mer
2082_b.q	5'-gTg CAg CAT ggT gAC CAC AgT TAA T-3'	25mer
Rep.p	5'-gCA gAT ggA TCA CAA gAT CAT gAg TTT-3'	27mer
Rep.q	5'-CCA gAg gTg ATA ACA gTC TTC CAg AAT-3'	27mer

7.3 Anhang C: Deletionen und Basenaustausche

7.3.1 Daten aus dem Vergleich der Klone 98L15 (GBF) und 90B5 (Jena)

7.3.1.1 Deletionen

Es sind jeweils 5 bp vor und nach der Deletion angeben, um die Position auf dem jeweiligen Klon genau Lokalisieren zu können. Der deletierte Bereich auf dem jeweiligen Klon ist mit „*“ angegeben und zusätzlich blau markiert.

DELETION 1:

```
P90B5:      731 TTTT*ACATG      740
              |||||
98L15:  27117 TTTT*ACATG  27127
```

7.3.2 Daten aus dem Vergleich der Klone 49E3 (GBF) und R73i15 (Jena)

7.3.2.1 Deletionen

Es sind jeweils 5 bp vor und nach der Deletion angeben, um die Position auf dem jeweiligen Klon genau Lokalisieren zu können. Der deletierte Bereich auf dem jeweiligen Klon ist mit „*“ angegeben und zusätzlich blau markiert.

Bei der „Deletion 4“ findet sich eine Besonderheit. Auf dem Klon 49E3 (GBF) findet sich von der Position 8181-8333 ein 152 bp lange Sequenz, an deren Stelle auf dem Klon R73i15 (RIKEN) von der Position 52909-54356 eine 1447 bp lange Sequenz zu finden ist. Die 152 bp lange Sequenz des Klons 49E3 ist an keiner Position mit der 1447 bp langen Sequenz des Klons R73i15 deckungsgleich. Aus diesem Grund wurden die jeweiligen Sequenzen schon zu Anfang der Abweichung gegenübergestellt, bis zu der Position, an der die Sequenz des 49E3 endet. Ab dieser Position ist die fehlende Sequenz des Klons 49E3 durch Sternchen dargestellt.

DELETION 1:

```
R73i15:  49951 GGAAC*ACACACAGT*AAAAC  49969
              |||||
49E3 :    5233 GGAAC*****AAAAC    5242
```

DELETION 2:

```

73I15:  50775  AAAAAAAGTAGA  50786
          |||||  |||||
49E3 :   6048  AAAAA**GTAGA   6057

```

DELETION 3:

```

73I15:  50798 TTTTTCCTTT 50807
          |||||  |||||
49E3 :   6069 TTTTTCCTTT 6079

```

DELETION 4:

73I15:	52904	ATATACTTATATATATTATATATAAATTATATATAAATATATAAATATATTATATATAAATTA	52963
49E3 :	8176	ATATAATTATAAATGATATATAAATCAATAATATATCATTAATAATATATCATATATAATA	8235
73I15:	52964	TATATAAATTTATAAATATATTATATATATAAATATATAATATATTTTATATATTATATA	53023
49E3 :	8236	ATTAATATAAATTAATAATATAAATTTAAATATAATATATAAATTATAAATATAATATATAAT	8295
73I15:	53024	ATTATATATAATATATAAATATATTATATATATAAATATATAATATATTTTATATAAATA	53083
49E3 :	8296	TATAATTATATTTTTATGTATTTTAAATAATTCATATA*****	8333
73I15:	53084	TATAATTATAAATATATAATATATTTTTATATATTATATATAAATTATAAATTATTATATA	53143
49E3 :	8333	*****	8333
73I15:	53144	TAATCAATTAATAATATATCATTAATAATAATATATCATATATAAATAATATAATTAATAA	53203
49E3 :	8333	*****	8333
73I15:	53204	TATAATTAAAAATATAATATATAATTATAAATATAGTATATAATTATAATTATATTTTTTA	53263
49E3 :	8333	*****	8333
73I15:	53264	TGTATTTTTTAAATAATTCGTTTTAATAAATGAATAGAATTCATCACCTAACTTCTAAAGC	53323
49E3 :	8333	*****	8333
73I15:	53324	TACTTCTCCCGATGAAAGTCCGGGCACTTAGGAGTCAACAGTGGTGTATTGAGTGTATGA	53383
49E3 :	8333	*****	8333
73I15:	53384	CTCTTGCTCACAGGTGCTTCTCTCTGTCCCACGGCAGGCGGCACCAACAGAAGGCTTCT	53443
49E3 :	8333	*****	8333
73I15:	53444	GCAAAGCTGAGAGCTCCTTTTACCCATGGATCAGGAGCAGAGTTGCAGCTACACTGTCA	53503
49E3 :	8333	*****	8333
73I15:	53504	TGGCTGGTTTGGATGTTAGCAGTTAGTTAGCTTTGCGTCTTCAGAGCAATTGTTCTTAG	53563
49E3 :	8333	*****	8333
73I15:	53564	CTACTTAAAAAAGAAAAGAAAAGAAAGGAAAATTTAAAATTTCTTCATTGATCATTC	53623
49E3 :	8333	*****	8333
73I15:	53624	GGAGCATGTTTAATTTCCATGTATTTGTACAATTCAAAAGTTCACCTTGTTGTTGACTT	53683
49E3 :	8333	*****	8333
73I15:	53684	CTAGTTTTTATTCATTGTGGTCAGGAAAGATACTCGATATAATTTCAATTATTTTAAATT	53743

Anhang

```

49E3 : 8333 ***** 8333
73I15: 53744 TGTGAGACTTGTTTTGTGGCCTAACATATGGTCTATTCTGGAGAATGTTTCTTGCTG 53803
49E3 : 8333 ***** 8333
73I15: 53804 ATGAAAGAATGTGTTTTTGCATCTGTTGGATGAAATGTTCTGTAAATGTCTCTTTGTTA 53863
49E3 : 8333 ***** 8333
73I15: 53864 GTCTAAAATGCAGTTTTTTGTGTTTTTTGTTTTTTTAGGAGAGTCTCACTCTGTTGCCAGG 53923
49E3 : 8333 ***** 8333
73I15: 53924 CTGGAGTGCAGTAGCATGATCTTGACTCACTGCAACCTCCACCTCCCAGGTTCAAGCAAT 53983
49E3 : 8333 ***** 8333
73I15: 53984 TCTCATGCCTCAGCCTCCCAAGTAGCTGGAATTACAGATGCCACCACCACACCAGCCA 54043
49E3 : 8333 ***** 8333
73I15: 54044 ATTTTTGTATTTTATAGTAGAGACAGTGTTTTTACCATGCTGGCCAGGCTGGTCTCAAATC 54103
49E3 : 8333 ***** 8333
73I15: 54104 CTGGCCTCAAGGGATCCGCCTGCCTTGGCCTCCCAAGGTGCTGGTACTATAGGTGTGAGC 54163
49E3 : 8333 ***** 8333
73I15: 54164 CACCGCATCCAGCCTAAAGTACAGTATAAATCTAATGTTTTGTGTTTTGATCTTCTATCTA 54223
49E3 : 8333 ***** 8333
73I15: 54224 GACAATCTGTTTCAGTGCTGAGAATGGGGTATTGAATTCCCAGCTATTGTATTGGCATCT 54283
49E3 : 8333 ***** 8333
73I15: 54284 ATATCTCCCTTTAGACCTAATAATATTTGGTTTATATTATATATCTCGGTGCTCCAGTGT 54343
49E3 : 8333 ***** 8333
73I15: 54344 TGGGTGCATATATGTGCA 54361
              |||||
49E3 : 8333 *****GTGCA 8338

```

DELETION 5:

```

73I15: 56747 TTTT*GCCCC 56756
              |||||
49E3 : 10724 TTTT*GCCCC 10734

```

DELETION 6:

```

73I15: 58891 TAAATAAATAAATAATT 58908
              |||||
49E3 : 12869 TAAAT*****TAATT 12878

```

DELETION 7:

```

73I15: 59238 TTTT*AGGCT 59248
              |||||
49E3 : 13208 TTTT*AGGCT 13217

```

DELETION 8:

73I15: 59371 AAAAA*****CATTA 59380
 ||||| |||||
 49E3 : 13340 AAAAAATAAAACATTA 13356

DELETION 9:

73I15: 59542 AAAAA*****GAAAG 59551
 ||||| |||||
 49E3 : 13518 AAAAAA*****GAAAG 13541

DELETION 10:

73I15: 60140 TTTT*GTGGA 60149
 ||||| |||||
 49E3 : 14130 TTTT*GTGGA 14140

DELETION 11:

73I15: 62108 TTTT*AGATG 62118
 ||||| |||||
 49E3 : 16099 TTTT*AGATG 16108

DELETION 12:

73I15: 63054 CACCA*****CCATC 63064
 ||||| |||||
 49E3 : 17044 CACCAATCATCACCATCACTATCACCACCATCACCA*CCATC 17085

DELETION 13:

73I15: 63749 TGGTT*CTGAGAATCACGTTTCCTC 63773
 ||||| |||||
 49E3 : 17770 TGGTT*****TCCTC 17779

DELETION 14:

73I15: 64861 AATCT**GATTT 64870
 ||||| |||||
 49E3 : 18867 AATCTCTGATTT 18878

DELETION 15:

73I15: 65240 AAAAA*CAATC 65253
 ||||| |||||
 49E3 : 19248 AAAAA*CAATC 19257

DELETION 16:

73I15: 65546 GGGGG*TAGGG 65556
 ||||| |||||
 49E3 : 19550 GGGGG*TAGGG 19559

DELETION 17:

73I15: 67478 AAAAA*GAAGG 67488
 ||||| |||||
 49E3 : 21481 AAAAA*GAAGG 21490

DELETION 18:

```

73I15: 68750 AATTT***TTTTT 68759
      |||||  |||||
49E3 : 22752 AATTTCTTTTTTT 22764

```

DELETION 19:

```

73I15: 71008 AAAAAAATAATATATATATATATATATATATATATATATATATAT 71043
      |||||  |||||
49E3 : 25013 AAAAAA*****TATAT 25022

```

DELETION 20:

```

73I15: 71553 AGGTT**GTGTG 71562
      |||||  |||||
49E3 : 25532 AGGTTGTGTGTG 25543

```

DELETION 21:

```

73I15: 82141 CTGACCTGCAA 82151
      |||||  |||||
49E3 : 36122 CTGAC*TGCAA 36131

```

DELETION 22:

```

73I15: 90539 TTTTTTGTGTT 90549
      |||||  |||||
49E3 : 44519 TTTTT*GTGTT 44528

```

DELETION 23:

```

73I15: 95622 CAGCCACCAGG 95632
      |||||  |||||
49E3 : 49601 CAGCC*CCAGG 49610

```

DELETION 24:

```

73I15: 106206 TTTTTTTAACTC 106217
      |||||  |||||
49E3 : 60184 TTTTT**AACTC 60193

```

DELETION 25:

```

73I15: 106449 GTTTTGTTGTGTT 106461
      |||||  |||||
49E3 : 60425 GTTTT***GTGTT 60434

```

DELETION 26:

```

73I15: 109937 TTTTA*****CAGTT 109946
      |||||  |||||
49E3 : 63910 TTTTAGACCCTTTAACAACATATCA 63938

```

DELETION 27:

```

73I15: 111602 TTTTTTTTTTGAGAT 111616
      |||||  |||||
47E3 : 65594 TTTTT*****GAGAT 65603

```

DELETION 28:

```

73I15: 116923 TATTTATTGAAAT 116937
          |||||      |||||
49E3 : 70910 TATTT***GAAAT 70920

```

DELETION 28:

```

73I15: 118983 AAAAA*TAAAT 118992
          |||||      |||||
49E3 : 72966 AAAAAATAAAT 72976

```

DELETIONEN 28 u. 30:

```

73I15: 126629 TACA*ATATAT 126639
          |||||      |||||
49E3 : 80613 TACAATTATATAT 80625

```

DELETION 31:

```

73I15: 127901 GACTGTATATAT 127912
          |||||      |||||
49E3 : 81887 GACTG**TATAT 81896

```

7.3.2.2 Einzelbasenaustausche

Klon	Position	Basen- austausch	Position	Basen- austausch	Position	Basen- austausch
73I15 (RIKEN)	45403	T	45477	G	45558	G
49E3 (GBF)	685	C	759	C	840	A
73I15 (RIKEN)	46322	T	46465	A	46653	C
49E3 (GBF)	1604	A	1747	G	1935	T
73I15 (RIKEN)	47056	A	47095	G	47850	T
49E3 (GBF)	2338	G	2377	C	3132	G
73I15 (RIKEN)	48117	G	48512	G	48926	T
49E3 (GBF)	3399	A	3794	C	4208	C
73I15 (RIKEN)	49120	C	49353	A	49370	G
49E3 (GBF)	4402	G	4635	G	4652	A
73I15 (RIKEN)	49495	C	49526	A	49578	A
49E3 (GBF)	4777	A	4808	T	4860	G
73I15 (RIKEN)	50187	T	50205	A	50225	T
49E3 (GBF)	5460	G	5478	C	5498	C
73I15 (RIKEN)	50332	A	50442	A	50931	C
49E3 (GBF)	5605	G	5715	G	6203	T
73I15 (RIKEN)	50948	G	51004	G	51181	C
49E3 (GBF)	6220	A	6276	A	6443	T

Anhang

Klon	Position	Basen- austausch	Position	Basen- austausch	Position	Basen- austausch
73I15 (RIKEN)	51565	T	51619	A	52567	T
49E3 (GBF)	6837	C	6891	C	7839	C
73I15 (RIKEN)	52633	G	54579	T	54844	C
49E3 (GBF)	7905	T	8556	A	8821	T
73I15 (RIKEN)	54917	G	54996	C	55537	G
49E3 (GBF)	8894	A	8973	T	9514	A
73I15 (RIKEN)	55565	G	55839	C	55892	G
49E3 (GBF)	9542	A	9816	A	9869	A
73I15 (RIKEN)	56182	C	56357	G	56448	T
49E3 (GBF)	10159	A	10334	C	10425	C
73I15 (RIKEN)	56932	G	57142	G	57153	G
49E3 (GBF)	10910	C	11120	A	11131	A
73I15 (RIKEN)	57268	G	57288	G	57393	G
49E3 (GBF)	11246	A	11266	A	11371	A
73I15 (RIKEN)	57521	T	57553	A	57560	T
49E3 (GBF)	11499	C	11531	G	11538	C
73I15 (RIKEN)	57601	T	57770	A	57808	C
49E3 (GBF)	11579	A	11748	G	11786	T
73I15 (RIKEN)	57880	A	57903	A	57923	C
49E3 (GBF)	11858	G	11881	G	11901	T
73I15 (RIKEN)	57960	T	57963	G	58029	G
49E3 (GBF)	11938	C	11941	A	12007	T
73I15 (RIKEN)	58048	G	58114	T	58125	C
49E3 (GBF)	12026	A	12092	G	12103	T
73I15 (RIKEN)	58151	C	58265	G	58471	T
49E3 (GBF)	12129	T	12243	A	12449	C
73I15 (RIKEN)	58565	C	58592	T	58708	A
49E3 (GBF)	12543	A	12570	C	12686	G
73I15 (RIKEN)	58884	A	58888	A	58892	A
49E3 (GBF)	12862	T	12866	T	12870	T
73I15 (RIKEN)	59014	C	59204	C	59455	G
49E3 (GBF)	12984	T	13174	G	13431	A
73I15 (RIKEN)	59654	T	59901	A	59920	G
49E3 (GBF)	13644	G	13891	T	13910	A

Anhang

Klon	Position	Basen- austausch	Position	Basen- austausch	Position	Basen- austausch
73I15 (RIKEN)	59939	T	60025	G	60062	A
49E3 (GBF)	13929	G	14015	A	14052	G
73I15 (RIKEN)	60112	T	60196	T	60430	G
49E3 (GBF)	14102	G	14187	C	14421	A
73I15 (RIKEN)	60498	A	60676	T	60893	G
49E3 (GBF)	14489	C	14667	C	14884	A
73I15 (RIKEN)	61068	C	61097	G	61134	G
49E3 (GBF)	15059	A	15088	T	15125	A
73I15 (RIKEN)	61206	T	61424	G	61674	A
49E3 (GBF)	15197	C	15415	A	15665	C
73I15 (RIKEN)	62020	G	62148	C	62347	T
49E3 (GBF)	16011	A	16138	T	16337	C
73I15 (RIKEN)	62631	C	64311	G	64540	A
49E3 (GBF)	16621	A	18317	T	18546	G
73I15 (RIKEN)	64564	C	64597	C	64630	G
49E3 (GBF)	18570	A	18603	A	18636	C
73I15 (RIKEN)	64975	G	65432	A	65455	C
49E3 (GBF)	18983	C	19436	G	19459	T
73I15 (RIKEN)	65606	C	66442	G	66645	T
49E3 (GBF)	19609	G	20445	A	20648	C
73I15 (RIKEN)	66667	C	66748	C	67008	C
49E3 (GBF)	20670	A	20751	T	21011	T
73I15 (RIKEN)	67010	T	67179	T	67212	T
49E3 (GBF)	21013	C	21182	C	21215	G
73I15 (RIKEN)	67429	A	67446	G	67651	C
49E3 (GBF)	21432	G	21449	A	21643	G
73I15 (RIKEN)	68536	G	68854	C	68868	T
49E3 (GBF)	22538	A	22859	T	22873	C
73I15 (RIKEN)	68944	A	69382	C	71063	A
49E3 (GBF)	22949	G	23387	T	25042	T
73I15 (RIKEN)	79933	C	80507	T	81191	C
49E3 (GBF)	33914	A	34488	C	35172	T

Anhang

Klon	Position	Basen- austausch	Position	Basen- austausch	Position	Basen- austausch
73I15 (RIKEN)	81869	T	82258	G	82335	T
49E3 (GBF)	35850	C	36238	C	36315	G
73I15 (RIKEN)	82452	A	83463	G	83519	T
49E3 (GBF)	36432	C	37443	A	37499	C
73I15 (RIKEN)	83742	T	83890	A	84265	G
49E3 (GBF)	37722	C	37870	G	38245	A
73I15 (RIKEN)	84363	A	84746	C	84817	A
49E3 (GBF)	38343	G	38726	T	38797	C
73I15 (RIKEN)	85616	T	85853	G	86042	C
49E3 (GBF)	39596	A	39833	A	40022	G
73I15 (RIKEN)	87113	A	89639	G	90398	C
49E3 (GBF)	41093	G	43619	A	44378	A
73I15 (RIKEN)	92139	C	92642	C	93813	A
49E3 (GBF)	46118	T	46621	T	47792	G
73I15 (RIKEN)	94014	A	94079	C	94430	T
49E3 (GBF)	47993	C	48058	G	48409	A
73I15 (RIKEN)	95154	G	99086	T	99538	A
49E3 (GBF)	49133	C	53064	A	53516	G
73I15 (RIKEN)	100128	T	100132	C	100825	C
49E3 (GBF)	54106	G	54110	T	54803	T
73I15 (RIKEN)	102662	A	104133	C	104435	A
49E3 (GBF)	56640	T	58111	T	58413	T
73I15 (RIKEN)	104575	G	105282	A	105448	C
49E3 (GBF)	58553	A	59260	G	59426	A
73I15 (RIKEN)	107531	T	108335	A	108504	C
49E3 (GBF)	61504	A	62308	G	62477	T
73I15 (RIKEN)	109468	T	110844	T	112624	T
49E3 (GBF)	63441	A	64836	C	66611	C
73I15 (RIKEN)	112757	C	114130	A	115266	A
49E3 (GBF)	66744	T	68117	G	69253	G
73I15 (RIKEN)	116263	C	116425	T	117335	A
49E3 (GBF)	70250	G	70412	C	71318	G

Anhang

Klon	Position	Basen- austausch	Position	Basen- austausch	Position	Basen- austausch
73I15 (RIKEN)	120109	A	120197	A	120308	T
49E3 (GBF)	74093	G	74181	C	74292	C
73I15 (RIKEN)	121050	A	123500	T	123978	G
49E3 (GBF)	75034	G	77484	C	77962	A
73I15 (RIKEN)	125806	G	126018	T	126095	C
49E3 (GBF)	79790	A	80002	A	80079	T
73I15 (RIKEN)	126169	A	126331	T	126349	C
49E3 (GBF)	80153	G	80315	C	80333	A
73I15 (RIKEN)	126475	G	126591	A	126634	A
49E3 (GBF)	80459	A	80575	G	80618	T
73I15 (RIKEN)	126647	A	126811	A	128663	G
49E3 (GBF)	80633	G	80797	G	82647	A
73I15 (RIKEN)	129040	T	130436	T	131673	C
49E3 (GBF)	83024	C	84420	C	85657	T
73I15 (RIKEN)	132106	C	132190	C	132258	A
49E3 (GBF)	86090	A	86174	G	86242	C
73I15 (RIKEN)	132309	G	132674	T	132792	C
49E3 (GBF)	86293	A	86658	C	86776	A
73I15 (RIKEN)	132814	C	133142	A	133328	C
49E3 (GBF)	86798	A	87127	C	87312	T
73I15 (RIKEN)	133387	G	133594	C	133787	A
49E3 (GBF)	87371	A	87578	T	87771	C
73I15 (RIKEN)	134092	C				
49E3 (GBF)	88076	T				

7.3.2.3 Mehrfache Basenaustausche

Klon	Anfangsposition	Endposition	Basenaustausche
73I15 (RIKEN)	67422	67423	CG
49E3 (GBF)	21425	21426	TA
73I15 (RIKEN)	94623	94625	AAA
49E3 (GBF)	48602	48604	TTT
73I15 (RIKEN)	133134	133135	CT
49E3 (GBF)	87118	87119	TG

7.4 Anhang D: Annotierungsdaten

Rohannotierungsdaten, die automatisch mit dem Programm „annomitter“ durch Vergleich der Sequenz mit verschiedenen Datenbanken erstellt wurden. Als Beispiel sind hier die Annotierungsdaten vom Dezember 2001 für den Klon 98L15 von Chromosom 21 aufgeführt. Die Sequenz des Klons, die normalerweise am Ende des Annotierungsergebnisses angehängt ist, wurde hier weggelassen.

```
ID   HS98L15   standard; DNA; HUM; 116189 BP.
XX
AC   ;
XX
DE   Homo sapiens chromosome 9 BAC 98L15.
XX
KW   HTG.
XX
OS   Homo sapiens (human)
OC   .
XX
RN   [1]
RP   1-116189
RA   Schoen O., Nordsiek G., Scharfe M., Conrad A., Hornischer K., Loehnert T.H.,
RA   Thies S., Bloecker H.;
RT   ;
RL   Submitted (DD-MMM-YYYY) to the EMBL/GenBank/DDBJ databases.
RL   GBF, Dept. of Genome Analysis, Mascheroder Weg 1,
RL   D-38124 Braunschweig, Germany,
RL   E-mail: info.genome@gbf.de
XX
CC   All annotations in this database entry are developed by
CC   computational tools. It is therefore not explicitly noted in the
CC   feature lines that evidence is not experimental.
CC   *
CC   ----- Genome Center
CC   Center: GBF, Braunschweig
CC   Center code: GBF
CC   Web site: http://genome.gbf.de/
CC   Contact: info.genome@gbf.de
CC   ----- Project Information
CC   Center project name:
CC   Center clone name: 98L15
CC   ----- Summary Statistics
CC   Sequencing vector: ###;
CC   Chemistry: Dye-terminator-amersham: ##% of reads
CC   Chemistry: Dye-primer-amersham: ##% of reads
CC   Assembly program: ##
CC   Consensus quality: 0 bases at least Q40
CC   Consensus quality: 0 bases at least Q30
CC   Consensus quality: 0 bases at least Q20
CC   Estimated insert size: ##; agarose-fp estimation
CC   Estimated insert size: 116189; sum-of-contigs estimation
CC   -----
CC   *
XX
CC   PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
CC   ++++++
CC   +   Analysis and annotation were performed with the automatic   +
CC   +   "first-pass" annotation and submission tool                 +
CC   +   "AnnoMitter" (Hornischer & Bloecker).                      +
CC   +   Programs used by "AnnoMitter":                             +
CC   ++++++
```

```

CC > GeneFinder (Green), Vers. 084
CC . Organism: human
CC > GenScan (Burge & Karlin), Vers. 1.0
CC . Used matrix: vertebrate; Minimum score: 0
CC > Grail (Xu et al.), Vers. 1.3
CC . Organism: human
CC > Mzef (Zhang)
CC . Prior probability: 0.04; Overlapping number: 0
CC > Xpound (Thomas & Skolnick)
CC . Base score cutoff: 0.2; Minimal exon length: 3 bp
CC > "Repeats": BLASTN 2.0.14 (Altschul et al.)
CC . Database(s): * RepBase: ALU (human), released 22-DEC-1995
CC . * RepBase: THR ((human), released 22-DEC-1995
CC . * RepBase: L1 (primate), released 22-DEC-1995
CC . * RepBase: MIR (primate), released 22-DEC-1995
CC . * RepBase: MER (primate), released 22-DEC-1995
CC . * RepBase: MIR2 (primate), released 22-DEC-1995
CC . * RepBase: THE (primate), released 22-DEC-1995
CC . Minimum score: 60; Minimum identity: 70 %;
CC > "ESTs": BLASTN 2.0.14 (Altschul et al.)
CC . Database(s): * embl (EST, human), released -DEC-
CC . * embl (EST, other), released -DEC-
CC . * emblnew (EST), Vers. 67+ (01-JAN-1970)
CC . Using sequence with masked repeats
CC . Minimum score: 60; Minimum identity: 70 %;
CC > "Tandem Repeats": GDE 2.2 option "tandem"
CC . Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
CC . Treat N's as mismatches? YES; Allow uniform consensi? NO
CC > "Inverted Repeats": GDE 2.2 option "inverted"
CC > "Micro Satellites": GDE 2.2 option "sputnik" (Abajian)
CC > "CpG Islands": GDE 2.2 option "cpg"
CC . CpG island region size 100 bp;
CC . Minimum GC contents 50 %; Observed/Expected 0.6
CC > "STS Scan": e-PCR (Schuler)
CC . Margin: 50; Number of mismatches allowed: 0; Word size: 7
CC . STS database: "dbSTS markers"
CC > "tRNA Scan": tRNAscan-SE (Lowe & Eddy), Vers. 1.11
XX
FH Key Location/Qualifiers
FH
FH source 1..116189
FH /organism="Homo sapiens"
FH /chromosome="9"
FH /clone="98L15"
FH exon complement(1..3)
FH /note="XPOUND prediction, score = 0.233"
FH repeat_region 1..957
FH /rpt_family="L1"
FH /note="96% identity: matches 2147..3103 of consensus"
FH misc_feature 1..116189
FH /note="assembly_fragment~clone_end:T7~vector_side:left"
FH /note="assembly_fragment~clone_end:SP6~vector_side:right"
FH exon 245..817
FH /note="GRAIL, score = 44%, comment = marginal"
FH satellite 405..418
FH /note="CACCA repeat"
FH repeat_region 1010..3696
FH /rpt_family="L1"
FH /note="96% identity: matches 3170..5858 of consensus"
FH exon 1138..3204
FH /note="GENSCAN prediction, score = 41.37"
FH exon 1357..1507
FH /note="GRAIL, score = 71%, comment = good"
FH exon 1713..2278
FH /note="MZEFE prediction, score = 0.746"
FH repeat_region 3723..3832
FH /rpt_family="L1"
FH /note="98% identity: matches 212..321 of consensus"
FH satellite 3846..3858
FH /note="ATA repeat"

```

```

FT    repeat_region    3851..3885
FT                                /rpt_type=INVERTED
FT                                /note="IR1, 85% complementary to IR1' (5518..5553)"
FT    exon              complement(4034..4187)
FT                                /note="XPOUND prediction, score = 0.701"
FT    exon              4058..4175
FT                                /note="GENSCAN prediction, score = 3.83"
FT    exon              complement(4069..4171)
FT                                /note="GRAIL, score = 63%, comment = good shadow"
FT    repeat_region    4846..4885
FT                                /rpt_type=TANDEM
FT                                /rpt_family="cttgattgggacaatgggtga repeat"
FT                                /note="homology = 100.0%, counts = 2"
FT    repeat_region    5518..5553
FT                                /rpt_type=INVERTED
FT                                /note="IR1', 85% complementary to IR1 (3851..3885)"
FT    satellite         5525..5552
FT                                /note="TTTAT repeat"
FT    repeat_region    5525..5554
FT                                /rpt_type=TANDEM
FT                                /rpt_family="tttat repeat"
FT                                /note="homology = 96.7%, counts = 6"
FT    repeat_region    complement(5561..5825)
FT                                /rpt_family="ALU"
FT                                /note="88% identity: matches 313..578 of consensus"
FT    repeat_region    5561..5827
FT                                /rpt_family="L1"
FT                                /note="86% identity: matches 250..518 of consensus"
FT    repeat_region    5572..5835
FT                                /rpt_type=INVERTED
FT                                /note="IR2, 85% complementary to IR2' (6881..7145)"
FT    exon              6818..6979
FT                                /note="MZEFP prediction, score = 0.541"
FT    repeat_region    6881..7145
FT                                /rpt_type=INVERTED
FT                                /note="IR2', 85% complementary to IR2 (5572..5835)"
FT    repeat_region    6881..7159
FT                                /rpt_family="ALU"
FT                                /note="94% identity: matches 1..281 of consensus"
FT    repeat_region    complement(6881..7159)
FT                                /rpt_family="L1"
FT                                /note="89% identity: matches 245..526 of consensus"
FT    misc_feature      6884..6990
FT                                /note="CpG_island (%GC=63.6, o/e=1.12, #CpGs=9)"
FT    exon              6948..6979
FT                                /note="XPOUND prediction, score = 0.289"
FT    repeat_region    7474..7599
FT                                /rpt_type=TANDEM
FT                                /rpt_family="taatatt repeat"
FT                                /note="homology = 60.3%, counts = 21"
FT    satellite         7756..7771
FT                                /note="TATT repeat"
FT    satellite         7775..7790
FT                                /note="TTA repeat"
FT    repeat_region    complement(7791..7899)
FT                                /rpt_family="L1"
FT                                /note="94% identity: matches 456..564 of consensus"
FT    exon              complement(9656..9784)
FT                                /note="GRAIL, score = 47%, comment = marginal"
FT    exon              10265..10295
FT                                /note="XPOUND prediction, score = 0.416"
FT    repeat_region    10563..11102
FT                                /rpt_family="MER11"
FT                                /note="92% identity: matches 1..538 of consensus"
FT    repeat_region    11084..11286
FT                                /rpt_family="MER11"
FT                                /note="87% identity: matches 532..732 of consensus"
FT    repeat_region    11251..11533
FT                                /rpt_family="MER11"
FT                                /note="88% identity: matches 814..1094 of consensus"

```

Anhang

```

FT     exon      complement(11667..11709)
FT                                     /note="XPOUND prediction, score = 0.397"
FT     exon      complement(12716..12843)
FT                                     /note="GRAIL, score = 88%, comment = excellent shadow"
FT     exon      complement(12735..12843)
FT                                     /note="XPOUND prediction, score = 0.578"
FT     exon      complement(12742..12843)
FT                                     /note="MZEFE prediction, score = 0.968"
FT     misc_feature complement(12771..12846)
FT                                     /note="match: EST BI460152"
FT     exon      complement(13038..13105)
FT                                     /note="MZEFE prediction, score = 0.947"
FT     repeat_region 13544..13713
FT                                     /rpt_family="L1"
FT                                     /note="100% identity: matches 4327..4496 of consensus"
FT     repeat_region 13714..13876
FT                                     /rpt_family="L1"
FT                                     /note="98% identity: matches 615..777 of consensus"
FT     repeat_region 13885..15460
FT                                     /rpt_family="L1"
FT                                     /note="94% identity: matches 557..2132 of consensus"
FT     exon      14038..15168
FT                                     /note="GRAIL, score = 65%, comment = good"
FT     exon      14395..14813
FT                                     /note="GENSCAN prediction, score = 8.45"
FT     exon      complement(15376..15419)
FT                                     /note="XPOUND prediction, score = 0.453"
FT     repeat_region 15449..15525
FT                                     /rpt_type=TANDEM
FT                                     /rpt_family="ttaaaataaaa repeat"
FT                                     /note="homology = 76.6%, counts = 7"
FT     satellite    15456..15468
FT                                     /note="ATA repeat"
FT     repeat_region 16580..16699
FT                                     /rpt_family="ALU"
FT                                     /note="87% identity: matches 1..120 of consensus"
FT     repeat_region complement(16580..16699)
FT                                     /rpt_family="L1"
FT                                     /note="84% identity: matches 477..596 of consensus"
FT     repeat_region complement(16741..16860)
FT                                     /rpt_family="L1"
FT                                     /note="85% identity: matches 167..286 of consensus"
FT     repeat_region 16741..16866
FT                                     /rpt_family="THR"
FT                                     /note="90% identity: matches 86..211 of consensus"
FT     misc_feature complement(16943..17054)
FT                                     /note="match: EST BI460152"
FT     exon      complement(16945..17016)
FT                                     /note="MZEFE prediction, score = 0.618"
FT     repeat_region 17621..17658
FT                                     /rpt_family="THE"
FT                                     /note="97% identity: matches 32..69 of consensus"
FT     exon      complement(17834..17884)
FT                                     /note="GRAIL, score = 83%, comment = excellent shadow"
FT     exon      complement(17839..17878)
FT                                     /note="XPOUND prediction, score = 0.619"
FT     repeat_region 19289..19542
FT                                     /rpt_family="ALU"
FT                                     /note="87% identity: matches 167..417 of consensus"
FT     repeat_region complement(19290..19542)
FT                                     /rpt_family="L1"
FT                                     /note="84% identity: matches 170..418 of consensus"
FT     exon      complement(20018..20403)
FT                                     /note="MZEFE prediction, score = 0.797"
FT     misc_feature complement(20018..20407)
FT                                     /note="match: EST BI460152"
FT     exon      complement(20051..20403)
FT                                     /note="GRAIL, score = 55%, comment = good"
FT     exon      complement(20134..20156)
FT                                     /note="XPOUND prediction, score = 0.208"

```

```

FT      exon      complement(20162..20205)
FT      /note="XPOUND prediction, score = 0.337"
FT      repeat_region 21135..21174
FT      /rpt_type=TANDEM
FT      /rpt_family="ctat repeat"
FT      /note="homology = 80.0%, counts = 10"
FT      exon      21396..21503
FT      /note="MZEF prediction, score = 0.854"
FT      exon      22025..22179
FT      /note="MZEF prediction, score = 0.924"
FT      repeat_region 22066..22133
FT      /rpt_family="ALU"
FT      /note="91% identity: matches 119..187 of consensus"
FT      repeat_region complement(22101..22360)
FT      /rpt_family="L1"
FT      /note="89% identity: matches 245..506 of consensus"
FT      repeat_region 22102..22360
FT      /rpt_family="ALU"
FT      /note="96% identity: matches 6..263 of consensus"
FT      misc_feature 22128..22318
FT      /note="CpG_island (%GC=60.7, o/e=1.06, #CpGs=17)"
FT      repeat_region 22361..22598
FT      /rpt_type=TANDEM
FT      /rpt_family="aaaataaaaataaaa repeat"
FT      /note="homology = 58.0%, counts = 17"
FT      exon      complement(22700..22740)
FT      /note="XPOUND prediction, score = 0.649"
FT      satellite   23056..23080
FT      /note="TTTTA repeat"
FT      exon      complement(23168..23181)
FT      /note="XPOUND prediction, score = 0.323"
FT      repeat_region 23178..23426
FT      /rpt_family="ALU"
FT      /note="87% identity: matches 468..716 of consensus"
FT      repeat_region complement(23185..23441)
FT      /rpt_family="L1"
FT      /note="84% identity: matches 167..420 of consensus"
FT      satellite   23449..23470
FT      /note="AAAC repeat"
FT      exon      24222..24359
FT      /note="MZEF prediction, score = 0.764"
FT      satellite   24878..24891
FT      /note="ATTA repeat"
FT      satellite   25111..25123
FT      /note="AC repeat"
FT      satellite   25826..25841
FT      /note="AAAAC repeat"
FT      exon      complement(26222..26225)
FT      /note="XPOUND prediction, score = 0.209"
FT      exon      complement(26231..26241)
FT      /note="XPOUND prediction, score = 0.213"
FT      exon      complement(26557..26582)
FT      /note="XPOUND prediction, score = 0.237"
FT      exon      complement(27527..27551)
FT      /note="XPOUND prediction, score = 0.342"
FT      repeat_region 27541..27621
FT      /rpt_family="L1"
FT      /note="85% identity: matches 83..163 of consensus"
FT      exon      complement(27667..27685)
FT      /note="XPOUND prediction, score = 0.487"
FT      repeat_region 27698..27793
FT      /rpt_family="ALU"
FT      /note="87% identity: matches 956..1051 of consensus"
FT      satellite   27810..27824
FT      /note="CAA repeat"
FT      repeat_region 28010..28039
FT      /rpt_type=TANDEM
FT      /rpt_family="ac repeat"
FT      /note="homology = 93.3%, counts = 15"
FT      satellite   28012..28034

```



```

FT      /note="AC repeat"
FT      repeat_region 28103..28192
FT      /rpt_type=TANDEM
FT      /rpt_family="ataaaa repeat"
FT      /note="homology = 64.4%, counts = 15"
FT      exon 28343..28525
FT      /note="GRAIL, score = 57%, comment = good"
FT      exon 28373..28509
FT      /note="MZEF prediction, score = 0.712"
FT      repeat_region 28585..28825
FT      /rpt_family="L1"
FT      /note="83% identity: matches 612..849 of consensus"
FT      repeat_region 28837..29124
FT      /rpt_family="L1"
FT      /note="82% identity: matches 1713..2001 of consensus"
FT      repeat_region 29022..29419
FT      /rpt_family="L1"
FT      /note="82% identity: matches 1..400 of consensus"
FT      repeat_region 29443..29493
FT      /rpt_family="L1"
FT      /note="96% identity: matches 2318..2368 of consensus"
FT      repeat_region 29509..29721
FT      /rpt_family="L1"
FT      /note="84% identity: matches 2146..2354 of consensus"
FT      repeat_region 29843..29949
FT      /rpt_family="L1"
FT      /note="87% identity: matches 8..115 of consensus"
FT      repeat_region 30012..30156
FT      /rpt_family="L1"
FT      /note="90% identity: matches 1130..1274 of consensus"
FT      repeat_region 30101..30249
FT      /rpt_family="L1"
FT      /note="81% identity: matches 23..173 of consensus"
FT      exon 30414..30564
FT      /note="GENSCAN prediction, score = 3.70"
FT      /note="GRAIL, score = 74%, comment = good"
FT      repeat_region 30421..30543
FT      /rpt_family="L1"
FT      /note="85% identity: matches 3310..3431 of consensus"
FT      repeat_region 30631..30761
FT      /rpt_family="L1"
FT      /note="87% identity: matches 25..156 of consensus"
FT      repeat_region 30709..30881
FT      /rpt_family="L1"
FT      /note="81% identity: matches 1721..1894 of consensus"
FT      repeat_region 30801..31520
FT      /rpt_family="L1"
FT      /note="84% identity: matches 3572..4290 of consensus"
FT      exon complement(30807..30909)
FT      /note="GRAIL, score = 85%, comment = excellent shadow"
FT      repeat_region 31379..31720
FT      /rpt_family="L1"
FT      /note="82% identity: matches 100..434 of consensus"
FT      repeat_region 31751..32371
FT      /rpt_family="L1"
FT      /note="81% identity: matches 129..758 of consensus"
FT      repeat_region 32304..32448
FT      /rpt_family="L1"
FT      /note="82% identity: matches 16..151 of consensus"
FT      repeat_region 32448..34716
FT      /rpt_family="L1"
FT      /note="94% identity: matches 3647..5919 of consensus"
FT      exon 32498..33085
FT      /note="GRAIL, score = 75%, comment = excellent"
FT      exon 33095..33246
FT      /note="GRAIL, score = 50%, comment = good"
FT      exon complement(33738..33772)
FT      /note="GRAIL, score = 54%, comment = good shadow"
FT      exon complement(34374..34439)
FT      /note="MZEF prediction, score = 0.863"

```

Anhang

FT	exon	34524..34672
FT		/note="GRAIL, score = 61%, comment = good"
FT	repeat_region	34744..34817
FT		/rpt_family="L1"
FT		/note="98% identity: matches 478..551 of consensus"
FT	satellite	34842..34854
FT		/note="AGAA repeat"
FT	repeat_region	34856..35062
FT		/rpt_family="L1"
FT		/note="85% identity: matches 178..384 of consensus"
FT	repeat_region	35089..35130
FT		/rpt_family="L1"
FT		/note="92% identity: matches 602..643 of consensus"
FT	repeat_region	35302..35373
FT		/rpt_type=TANDEM
FT		/rpt_family="aaaaagga repeat"
FT		/note="homology = 72.2%, counts = 9"
FT	satellite	35995..36005
FT		/note="AT repeat"
FT	exon	37121..37192
FT		/note="MZEF prediction, score = 0.919"
FT	exon	37121..37196
FT		/note="GRAIL, score = 84%, comment = excellent"
FT	exon	37777..37841
FT		/note="MZEF prediction, score = 0.826"
FT	misc_feature	complement(37795..37934)
FT		/note="match: EST BI460152"
FT	exon	37963..38076
FT		/note="MZEF prediction, score = 0.617"
FT	misc_feature	38235..38313
FT		/note="match: EST BI546167"
FT	misc_feature	complement(38263..38328)
FT		/note="match: EST BI492401"
FT	exon	38434..38519
FT		/note="GRAIL, score = 63%, comment = good"
FT	exon	38702..38874
FT		/note="GRAIL, score = 74%, comment = good"
FT	exon	complement(39216..39300)
FT		/note="MZEF prediction, score = 0.642"
FT	exon	complement(39225..39320)
FT		/note="GRAIL, score = 71%, comment = good shadow"
FT	repeat_region	complement(39764..39861)
FT		/rpt_family="L1"
FT		/note="84% identity: matches 829..926 of consensus"
FT	repeat_region	39766..39903
FT		/rpt_family="ALU"
FT		/note="83% identity: matches 402..539 of consensus"
FT	repeat_region	39881..40020
FT		/rpt_family="ALU"
FT		/note="80% identity: matches 131..270 of consensus"
FT	repeat_region	41143..41173
FT		/rpt_type=INVERTED
FT		/note="IR3, 80% complementary to IR3' (41176..41206)"
FT	repeat_region	41176..41206
FT		/rpt_type=INVERTED
FT		/note="IR3', 80% complementary to IR3 (41143..41173)"
FT	satellite	41607..41617
FT		/note="CA repeat"
FT	exon	41788..41883
FT		/note="MZEF prediction, score = 0.819"
FT	satellite	43404..43430
FT		/note="ATG repeat"
FT	satellite	43950..43973
FT		/note="AT repeat"
FT	repeat_region	43950..43999
FT		/rpt_type=TANDEM
FT		/rpt_family="at repeat"
FT		/note="homology = 78.0%, counts = 25"
FT	satellite	43988..43999
FT		/note="AT repeat"

```

FT   repeat_region   44080..44239
FT                               /rpt_type=TANDEM
FT                               /rpt_family="aatttttaaaaaaatt repeat"
FT                               /note="homology = 61.2%, counts = 10"
FT   repeat_region   44267..44322
FT                               /rpt_type=TANDEM
FT                               /rpt_family="cctt repeat"
FT                               /note="homology = 92.9%, counts = 14"
FT   satellite       44277..44323
FT                               /note="TTCC repeat"
FT   repeat_region   44678..44955
FT                               /rpt_family="ALU"
FT                               /note="89% identity: matches 5..282 of consensus"
FT   repeat_region   complement(44685..44937)
FT                               /rpt_family="L1"
FT                               /note="84% identity: matches 262..515 of consensus"
FT   repeat_region   45000..45063
FT                               /rpt_type=TANDEM
FT                               /rpt_family="at repeat"
FT                               /note="homology = 71.9%, counts = 32"
FT   exon            complement(46551..46701)
FT                               /note="GRAIL, score = 69%, comment = good"
FT   misc_feature     46974..47572
FT                               /note="CpG_island (%GC=67.3, o/e=0.88, #CpGs=53)"
FT   exon            46983..47267
FT                               /note="XPOUND prediction, score = 0.800"
FT   exon            47051..47267
FT                               /note="GENSCAN prediction, score = 21.30"
FT   misc_feature     (47145.47163)..47268
FT                               /note="match: ESTs AL596989 BI603848 BI753557"
FT   exon            47306..47416
FT                               /note="XPOUND prediction, score = 0.295"
FT   exon            48197..48297
FT                               /note="XPOUND prediction, score = 0.463"
FT   exon            49009..49110
FT                               /note="GRAIL, score = 75%, comment = excellent"
FT   satellite       49538..49551
FT                               /note="GAAGG repeat"
FT   misc_feature     50546..50711
FT                               /note="match: ESTs AL596989 BI753557"
FT   exon            50548..50710
FT                               /note="GRAIL, score = 60%, comment = good"
FT                               /note="MZEF prediction, score = 0.931"
FT   misc_feature     (52080.52081)..(52155.52334)
FT                               /note="match: ESTs AL596989 BI603848 BI753557"
FT   exon            53094..53120
FT                               /note="XPOUND prediction, score = 0.570"
FT   repeat_region   53259..53316
FT                               /rpt_type=TANDEM
FT                               /rpt_family="ta repeat"
FT                               /note="homology = 72.4%, counts = 29"
FT   repeat_region   53629..53666
FT                               /rpt_type=INVERTED
FT                               /note="IR4, 86% complementary to IR4' (53704..53741)"
FT   repeat_region   53704..53741
FT                               /rpt_type=INVERTED
FT                               /note="IR4', 86% complementary to IR4 (53629..53666)"
FT   exon            54001..54017
FT                               /note="XPOUND prediction, score = 0.724"
FT   exon            complement(54908..55002)
FT                               /note="GRAIL, score = 52%, comment = good shadow"
FT   repeat_region   55007..55028
FT                               /rpt_type=INVERTED
FT                               /note="IR5, 90% complementary to IR5' (55043..55064)"
FT   repeat_region   55043..55064
FT                               /rpt_type=INVERTED
FT                               /note="IR5', 90% complementary to IR5 (55007..55028)"
FT   repeat_region   complement(55168..55224)
FT                               /rpt_family="THE"
FT                               /note="91% identity: matches 4..60 of consensus"

```

```

FT    misc_feature    55421..55522
FT                                     /note="match: ESTs BI753557 BI603848"
FT    exon            55634..55767
FT                                     /note="GRAIL, score = 91%, comment = excellent"
FT    exon            55707..55767
FT                                     /note="XPOUND prediction, score = 0.685"
FT    repeat_region   55825..56049
FT                                     /rpt_family="ALU"
FT                                     /note="81% identity: matches 856..1085 of consensus"
FT    repeat_region   complement(55878..56015)
FT                                     /rpt_family="L1"
FT                                     /note="81% identity: matches 742..879 of consensus"
FT    exon            56288..56405
FT                                     /note="GRAIL, score = 86%, comment = excellent"
FT    repeat_region   57323..57604
FT                                     /rpt_family="ALU"
FT                                     /note="87% identity: matches 1..282 of consensus"
FT    repeat_region   complement(57336..57557)
FT                                     /rpt_family="L1"
FT                                     /note="86% identity: matches 207..424 of consensus"
FT    exon            57844..58278
FT                                     /note="GENSCAN prediction, score = 10.60"
FT                                     /note="GRAIL, score = 78%, comment = excellent"
FT    misc_feature    (57864.58036)..(58142.58265)
FT                                     /note="match: ESTs BI753557 BI603848"
FT    repeat_region   59475..59528
FT                                     /rpt_family="THE"
FT                                     /note="94% identity: matches 1..54 of consensus"
FT    exon            complement(59677..59786)
FT                                     /note="GRAIL, score = 67%, comment = good shadow"
FT    repeat_region   59682..59829
FT                                     /rpt_family="THE"
FT                                     /note="83% identity: matches 1..147 of consensus"
FT    repeat_region   59882..60654
FT                                     /rpt_family="THR"
FT                                     /note="83% identity: matches 47..824 of consensus"
FT    exon            60055..60143
FT                                     /note="XPOUND prediction, score = 0.648"
FT    exon            complement(60110..60138)
FT                                     /note="XPOUND prediction, score = 0.239"
FT    exon            60497..60940
FT                                     /note="GENSCAN prediction, score = 9.19"
FT    repeat_region   60692..61435
FT                                     /rpt_family="THR"
FT                                     /note="82% identity: matches 719..1460 of consensus"
FT    exon            complement(60796..60880)
FT                                     /note="GRAIL, score = 45%, comment = marginal"
FT    exon            60990..61035
FT                                     /note="XPOUND prediction, score = 0.215"
FT    exon            complement(61120..61170)
FT                                     /note="GRAIL, score = 49%, comment = marginal"
FT    repeat_region   61496..61549
FT                                     /rpt_family="THE"
FT                                     /note="94% identity: matches 1..54 of consensus"
FT    repeat_region   61583..61662
FT                                     /rpt_family="THE"
FT                                     /note="90% identity: matches 85..165 of consensus"
FT    exon            complement(61664..61763)
FT                                     /note="GRAIL, score = 66%, comment = good"
FT    exon            complement(61672..61763)
FT                                     /note="XPOUND prediction, score = 0.327"
FT    repeat_region   61745..61870
FT                                     /rpt_family="THE"
FT                                     /note="89% identity: matches 97..222 of consensus"
FT    exon            complement(62402..62470)
FT                                     /note="MZEFP prediction, score = 0.557"
FT    exon            complement(63290..63405)
FT                                     /note="GRAIL, score = 59%, comment = good"
FT    exon            63707..63729
FT                                     /note="XPOUND prediction, score = 0.233"

```

FT	exon	complement(63747..63795)
FT		/note="XPOUND prediction, score = 0.300"
FT	exon	complement(64674..64712)
FT		/note="GRAIL, score = 89%, comment = excellent"
FT	satellite	64714..64727
FT		/note="TATAA repeat"
FT	exon	complement(64972..65058)
FT		/note="MZEf prediction, score = 0.521"
FT	exon	65350..65455
FT		/note="MZEf prediction, score = 0.767"
FT	exon	complement(65730..65790)
FT		/note="XPOUND prediction, score = 0.287"
FT	exon	65848..65917
FT		/note="XPOUND prediction, score = 0.255"
FT	repeat_region	complement(66054..66164)
FT		/rpt_family="L1"
FT		/note="81% identity: matches 478..588 of consensus"
FT	repeat_region	66057..66153
FT		/rpt_family="ALU"
FT		/note="88% identity: matches 637..733 of consensus"
FT	repeat_region	66187..66311
FT		/rpt_family="ALU"
FT		/note="89% identity: matches 883..1007 of consensus"
FT	repeat_region	complement(66201..66319)
FT		/rpt_family="L1"
FT		/note="84% identity: matches 170..288 of consensus"
FT	exon	66259..66284
FT		/note="XPOUND prediction, score = 0.256"
FT	exon	66290..66309
FT		/note="XPOUND prediction, score = 0.208"
FT	satellite	66337..66353
FT		/note="AG repeat"
FT	exon	67191..67291
FT		/note="MZEf prediction, score = 0.880"
FT	exon	complement(67318..67411)
FT		/note="MZEf prediction, score = 0.752"
FT	satellite	68811..68837
FT		/note="TGAA repeat"
FT	repeat_region	68811..68838
FT		/rpt_type=TANDEM
FT		/rpt_family="tgaa repeat"
FT		/note="homology = 96.4%, counts = 7"
FT	repeat_region	69383..69454
FT		/rpt_type=TANDEM
FT		/rpt_family="gcactattagaaaatat repeat"
FT		/note="homology = 77.8%, counts = 4"
FT	exon	complement(72491..72541)
FT		/note="GRAIL, score = 71%, comment = good"
FT	misc_feature	72520..72573
FT		/note="match: EST BI535830"
FT	repeat_region	73169..73208
FT		/rpt_family="L1"
FT		/note="95% identity: matches 9..48 of consensus"
FT	repeat_region	73303..73384
FT		/rpt_family="L1"
FT		/note="89% identity: matches 51..132 of consensus"
FT	repeat_region	73666..73724
FT		/rpt_family="L1"
FT		/note="89% identity: matches 82..140 of consensus"
FT	exon	complement(73838..74105)
FT		/note="GENSCAN prediction, score = 3.88"
FT	exon	complement(73982..74082)
FT		/note="GRAIL, score = 67%, comment = good"
FT	repeat_region	74210..74247
FT		/rpt_type=TANDEM
FT		/rpt_family="gt repeat"
FT		/note="homology = 100.0%, counts = 19"
FT	satellite	74210..74247
FT		/note="GT repeat"
FT	exon	74271..74285

```

FT          /note="XPOUND prediction, score = 0.257"
FT exon      74450..74482
FT          /note="XPOUND prediction, score = 0.279"
FT exon      complement(74515..74582)
FT          /note="XPOUND prediction, score = 0.380"
FT exon      74771..74923
FT          /note="GRAIL, score = 48%, comment = marginal"
FT exon      complement(75359..75371)
FT          /note="XPOUND prediction, score = 0.270"
FT satellite 75935..75945
FT          /note="TG repeat"
FT exon      76013..76129
FT          /note="MZEFP prediction, score = 0.642"
FT exon      complement(76850..76875)
FT          /note="XPOUND prediction, score = 0.202"
FT exon      77969..78021
FT          /note="GRAIL, score = 93%, comment = excellent"
FT repeat_region complement(78585..78647)
FT          /rpt_family="THE"
FT          /note="92% identity: matches 4..66 of consensus"
FT repeat_region complement(78620..78739)
FT          /rpt_family="THE"
FT          /note="85% identity: matches 32..146 of consensus"
FT satellite 78775..78786
FT          /note="AG repeat"
FT repeat_region complement(78891..78969)
FT          /rpt_family="THE"
FT          /note="88% identity: matches 14..92 of consensus"
FT exon      complement(79136..79220)
FT          /note="GRAIL, score = 73%, comment = good"
FT exon      complement(79159..79318)
FT          /note="MZEFP prediction, score = 0.511"
FT exon      complement(79472..79554)
FT          /note="MZEFP prediction, score = 0.574"
FT repeat_region 79962..80112
FT          /rpt_family="MER11"
FT          /note="87% identity: matches 8..160 of consensus"
FT repeat_region 80150..80360
FT          /rpt_type=INVERTED
FT          /note="IR6, 95% complementary to IR6' (81395..81605)"
FT repeat_region 80363..80408
FT          /rpt_family="MER11"
FT          /note="95% identity: matches 212..257 of consensus"
FT repeat_region 80409..80772
FT          /rpt_family="MER11"
FT          /note="88% identity: matches 124..483 of consensus"
FT repeat_region 80786..80927
FT          /rpt_family="MER11"
FT          /note="86% identity: matches 603..744 of consensus"
FT repeat_region 80927..81021
FT          /rpt_family="MER11"
FT          /note="87% identity: matches 718..812 of consensus"
FT repeat_region 81023..81284
FT          /rpt_family="MER11"
FT          /note="86% identity: matches 653..908 of consensus"
FT repeat_region 81395..81605
FT          /rpt_type=INVERTED
FT          /note="IR6', 95% complementary to IR6 (80150..80360)"
FT repeat_region 81834..81985
FT          /rpt_family="L1"
FT          /note="83% identity: matches 693..848 of consensus"
FT repeat_region complement(81837..81984)
FT          /rpt_family="ALU"
FT          /note="90% identity: matches 234..381 of consensus"
FT satellite 83332..83347
FT          /note="CAAA repeat"
FT repeat_region 83643..83715
FT          /rpt_family="MER7"
FT          /note="87% identity: matches 21..93 of consensus"
FT repeat_region 83683..83833

```

```

FT          /rpt_family="MER7"
FT          /note="83% identity: matches 70..220 of consensus"
FT  satellite 84386..84399
FT          /note="AAC repeat"
FT  exon      84906..85045
FT          /note="GRAIL, score = 89%, comment = excellent"
FT  STS       84918..85023
FT          /standard_name="SHGC-74043 (D11S4975), Map: 21, Homo
FT          sapiens"
FT          /db_xref="dbSTS:71194"
FT          /note="GenBank Accession Number: G48499"
FT  exon      complement(84991..85043)
FT          /note="XPOUND prediction, score = 0.656"
FT  repeat_region 85053..85104
FT          /rpt_type=TANDEM
FT          /rpt_family="gt repeat"
FT          /note="homology = 98.1%, counts = 26"
FT  satellite 85055..85104
FT          /note="GT repeat"
FT  exon      85190..85246
FT          /note="GRAIL, score = 46%, comment = marginal"
FT  repeat_region 85479..85519
FT          /rpt_type=INVERTED
FT          /note="IR7, 78% complementary to IR7' (85556..85596)"
FT  repeat_region 85556..85596
FT          /rpt_type=INVERTED
FT          /note="IR7', 78% complementary to IR7 (85479..85519)"
FT  satellite 85557..85577
FT          /note="AT repeat"
FT  exon      complement(85655..85667)
FT          /note="XPOUND prediction, score = 0.280"
FT  exon      complement(85822..85849)
FT          /note="XPOUND prediction, score = 0.346"
FT  repeat_region complement(86442..87060)
FT          /rpt_family="L1"
FT          /note="85% identity: matches 75..693 of consensus"
FT  exon      complement(86652..86761)
FT          /note="XPOUND prediction, score = 0.630"
FT  exon      86760..86882
FT          /note="GRAIL, score = 60%, comment = good"
FT  exon      86987..87086
FT          /note="GRAIL, score = 63%, comment = good"
FT  repeat_region complement(87106..87181)
FT          /rpt_family="L1"
FT          /note="92% identity: matches 1168..1243 of consensus"
FT  exon      complement(87360..87405)
FT          /note="GRAIL, score = 85%, comment = excellent shadow"
FT  misc_feature 87500..87729
FT          /note="CpG_island (%GC=61.7, o/e=0.79, #CpGs=19)"
FT  repeat_region complement(87500..87769)
FT          /rpt_family="ALU"
FT          /note="96% identity: matches 8..278 of consensus"
FT  repeat_region 87500..87769
FT          /rpt_family="L1"
FT          /note="91% identity: matches 249..519 of consensus"
FT  repeat_region complement(87994..88052)
FT          /rpt_family="L1"
FT          /note="91% identity: matches 13..71 of consensus"
FT  repeat_region 88060..88116
FT          /rpt_type=INVERTED
FT          /note="IR8, 76% complementary to IR8' (88245..88300)"
FT  repeat_region 88128..88201
FT          /rpt_type=TANDEM
FT          /rpt_family="ta repeat"
FT          /note="homology = 66.2%, counts = 37"
FT  repeat_region 88203..88288
FT          /rpt_type=TANDEM
FT          /rpt_family="ta repeat"
FT          /note="homology = 66.3%, counts = 43"
FT  repeat_region 88245..88300

```

```

FT          /rpt_type=INVERTED
FT          /note="IR8', 76% complementary to IR8 (88060..88116)"
FT  exon    88511..88690
FT          /note="GRAIL, score = 41%, comment = marginal"
FT  repeat_region complement(88969..89200)
FT          /rpt_family="ALU"
FT          /note="85% identity: matches 14..244 of consensus"
FT  repeat_region 88969..89202
FT          /rpt_family="L1"
FT          /note="85% identity: matches 195..424 of consensus"
FT  repeat_region 88987..89053
FT          /rpt_type=INVERTED
FT          /note="IR9, 76% complementary to IR9' (89686..89753)"
FT  exon    complement(89148..89258)
FT          /note="MZEFP prediction, score = 0.613"
FT  repeat_region 89623..89673
FT          /rpt_family="ALU"
FT          /note="90% identity: matches 111..161 of consensus"
FT  repeat_region 89686..89753
FT          /rpt_type=INVERTED
FT          /note="IR9', 76% complementary to IR9 (88987..89053)"
FT  repeat_region 89689..89790
FT          /rpt_family="L1"
FT          /note="82% identity: matches 341..442 of consensus"
FT  exon    complement(89719..89728)
FT          /note="XPOUND prediction, score = 0.532"
FT  satellite 89808..89826
FT          /note="CAAAA repeat"
FT  exon    complement(93200..93239)
FT          /note="GRAIL, score = 84%, comment = excellent"
FT  exon    complement(96512..96677)
FT          /note="GENSCAN prediction, score = 5.40"
FT  repeat_region 96647..96688
FT          /rpt_family="THE"
FT          /note="95% identity: matches 42..83 of consensus"
FT  exon    complement(96788..96860)
FT          /note="GENSCAN prediction, score = 3.68"
FT  exon    complement(96788..96873)
FT          /note="XPOUND prediction, score = 0.244"
FT  repeat_region 96836..96994
FT          /rpt_family="THE"
FT          /note="83% identity: matches 47..207 of consensus"
FT  exon    complement(98610..98613)
FT          /note="XPOUND prediction, score = 0.200"
FT  exon    complement(98649..98665)
FT          /note="XPOUND prediction, score = 0.235"
FT  exon    98742..98893
FT          /note="GRAIL, score = 56%, comment = good"
FT  repeat_region 99015..99051
FT          /rpt_type=INVERTED
FT          /note="IR10, 78% complementary to IR10' (99738..99774)"
FT  repeat_region 99019..99042
FT          /rpt_type=TANDEM
FT          /rpt_family="at repeat"
FT          /note="homology = 100.0%, counts = 12"
FT  satellite 99019..99043
FT          /note="AT repeat"
FT  repeat_region 99108..99161
FT          /rpt_type=TANDEM
FT          /rpt_family="ttgacatttttaataat repeat"
FT          /note="homology = 85.2%, counts = 3"
FT  satellite 99204..99214
FT          /note="AT repeat"
FT  repeat_region 99733..99814
FT          /rpt_type=TANDEM
FT          /rpt_family="at repeat"
FT          /note="homology = 63.4%, counts = 41"
FT  repeat_region 99738..99774
FT          /rpt_type=INVERTED
FT          /note="IR10', 78% complementary to IR10 (99015..99051)"

```


Anhang

```

FT     exon          complement(99888..99903)
FT                                     /note="XPOUND prediction, score = 0.315"
FT     repeat_region complement(99917..100197)
FT                                     /rpt_family="L1"
FT                                     /note="83% identity: matches 245..528 of consensus"
FT     exon          complement(99918..99940)
FT                                     /note="XPOUND prediction, score = 0.215"
FT     repeat_region 99919..100218
FT                                     /rpt_type=INVERTED
FT                                     /note="IR11, 83% complementary to IR11' (100869..101169)"
FT     repeat_region 99924..100197
FT                                     /rpt_family="ALU"
FT                                     /note="86% identity: matches 153..426 of consensus"
FT     repeat_region 100198..100245
FT                                     /rpt_type=TANDEM
FT                                     /rpt_family="aaaaaaaaagaa repeat"
FT                                     /note="homology = 87.5%, counts = 4"
FT     exon          complement(100366..100382)
FT                                     /note="XPOUND prediction, score = 0.360"
FT     repeat_region 100869..101169
FT                                     /rpt_type=INVERTED
FT                                     /note="IR11', 83% complementary to IR11 (99919..100218)"
FT     repeat_region complement(100893..101169)
FT                                     /rpt_family="ALU"
FT                                     /note="86% identity: matches 1..277 of consensus"
FT     repeat_region 100895..101155
FT                                     /rpt_family="L1"
FT                                     /note="84% identity: matches 166..424 of consensus"
FT     exon          101573..101593
FT                                     /note="XPOUND prediction, score = 0.389"
FT     satellite      101930..101942
FT                                     /note="TG repeat"
FT     repeat_region 101930..101993
FT                                     /rpt_type=TANDEM
FT                                     /rpt_family="ta repeat"
FT                                     /note="homology = 76.6%, counts = 32"
FT     repeat_region 101940..101965
FT                                     /rpt_type=INVERTED
FT                                     /note="IR12, 96% complementary to IR12' (101976..102000)"
FT     satellite      101943..101961
FT                                     /note="AT repeat"
FT     satellite      101968..101982
FT                                     /note="TG repeat"
FT     repeat_region 101976..102000
FT                                     /rpt_type=INVERTED
FT                                     /note="IR12', 96% complementary to IR12 (101940..101965)"
FT     satellite      101983..101993
FT                                     /note="AT repeat"
FT     satellite      102064..102076
FT                                     /note="AT repeat"
FT     exon          complement(102451..102490)
FT                                     /note="MZEFP prediction, score = 0.566"
FT     exon          complement(102756..102825)
FT                                     /note="XPOUND prediction, score = 0.537"
FT     exon          102803..102882
FT                                     /note="GRAIL, score = 78%, comment = excellent"
FT     STS           103435..103556
FT                                     /standard_name="SHGC-52079 (D11S4945), Map: 21, Homo
FT                                     sapiens"
FT                                     /db_xref="dbSTS:51782"
FT                                     /note="GenBank Accession Number: G34791"
FT     STS           103459..103611
FT                                     /standard_name="WI-5366 (DXYS5Y), Map: 972_H_10, Homo
FT                                     sapiens"
FT                                     /db_xref="dbSTS:10390"
FT                                     /note="GenBank Accession Number: G04904"
FT     exon          complement(104802..104821)
FT                                     /note="XPOUND prediction, score = 0.329"
FT     misc_feature  104899..105129
FT                                     /note="CpG_island (%GC=70.1, o/e=0.72, #CpGs=23)"

```

Anhang

```

FT     exon          complement(104988..105077)
FT                                     /note="GENSCAN prediction, score = 15.54"
FT     exon          104989..105141
FT                                     /note="XPOUND prediction, score = 0.758"
FT     exon          105076..105141
FT                                     /note="GRAIL, score = 41%, comment = marginal"
FT     repeat_region 106334..106381
FT                                     /rpt_family="ALU"
FT                                     /note="95% identity: matches 146..193 of consensus"
FT     misc_feature   (106405.106419)..106476
FT                                     /note="match: ESTs BI711722 AL598672"
FT     misc_feature   complement((106418.106419)..(106476.106519))
FT                                     /note="match: ESTs BI711309 BI602470"
FT     repeat_region 107419..107682
FT                                     /rpt_family="ALU"
FT                                     /note="85% identity: matches 12..275 of consensus"
FT     repeat_region complement(107497..107685)
FT                                     /rpt_family="L1"
FT                                     /note="85% identity: matches 164..352 of consensus"
FT     repeat_region 107689..107705
FT                                     /rpt_type=INVERTED
FT                                     /note="IR13, 100% complementary to IR13' (108445..108461)"
FT     repeat_region 108445..108461
FT                                     /rpt_type=INVERTED
FT                                     /note="IR13', 100% complementary to IR13 (107689..107705)"
FT     repeat_region 108794..108824
FT                                     /rpt_type=INVERTED
FT                                     /note="IR14, 90% complementary to IR14' (109472..109501)"
FT     repeat_region 109472..109501
FT                                     /rpt_type=INVERTED
FT                                     /note="IR14', 90% complementary to IR14 (108794..108824)"
FT     repeat_region 110080..110112
FT                                     /rpt_type=INVERTED
FT                                     /note="IR15, 84% complementary to IR15' (110217..110249)"
FT     repeat_region 110217..110249
FT                                     /rpt_type=INVERTED
FT                                     /note="IR15', 84% complementary to IR15 (110080..110112)"
FT     repeat_region 110240..110283
FT                                     /rpt_type=TANDEM
FT                                     /rpt_family="ta repeat"
FT                                     /note="homology = 75.0%, counts = 22"
FT     satellite      111160..111172
FT                                     /note="AC repeat"
FT     exon          111697..111706
FT                                     /note="XPOUND prediction, score = 0.234"
FT     misc_feature   111921..111976
FT                                     /note="match: ESTs BI599980 BI602386"
FT     exon          complement(111950..111982)
FT                                     /note="GRAIL, score = 87%, comment = excellent"
FT     misc_feature   112043..112153
FT                                     /note="CpG_island (%GC=54.1, o/e=1.29, #CpGs=9)"
FT     misc_feature   (112079.112158)..(112203.112216)
FT                                     /note="match: ESTs BI461235 BI467347 BI561506"
FT     misc_feature   complement((112099.112771)..(112203.112896))
FT                                     /note="match: ESTs BI546404 BI758163 BI438420 BI495948
FT                                     BI753153 BI517778"
FT     misc_feature   (112438.112778)..(112592.112896)
FT                                     /note="match: ESTs AL601602 BI491171 BI495949 BI491261
FT                                     BI262258 BI467347 BI325030 BI521622 AL597539 BI252949"
FT     repeat_region complement(113641..113754)
FT                                     /rpt_family="L1"
FT                                     /note="91% identity: matches 413..526 of consensus"
FT     repeat_region 113641..113755
FT                                     /rpt_family="ALU"
FT                                     /note="93% identity: matches 1..115 of consensus"
FT     repeat_region 113681..113953
FT                                     /rpt_type=INVERTED
FT                                     /note="IR16, 86% complementary to IR16' (114871..115141)"
FT     repeat_region 113778..113924
FT                                     /rpt_family="ALU"

```

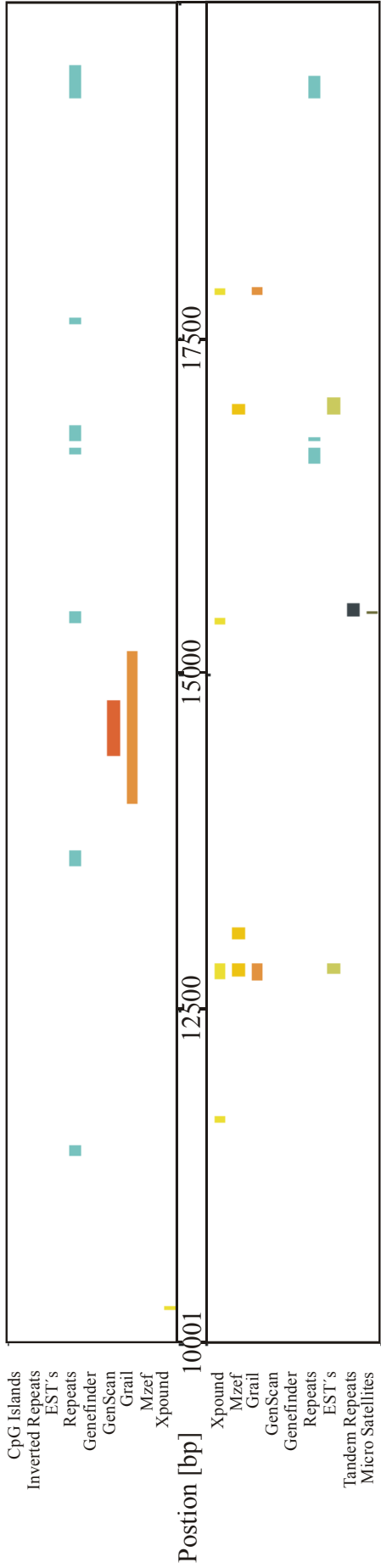
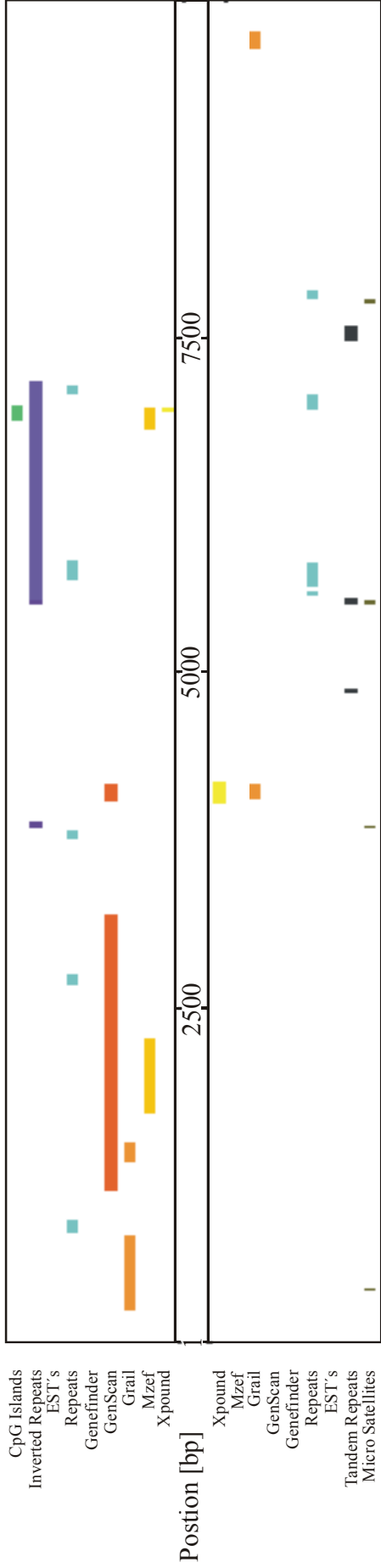
```

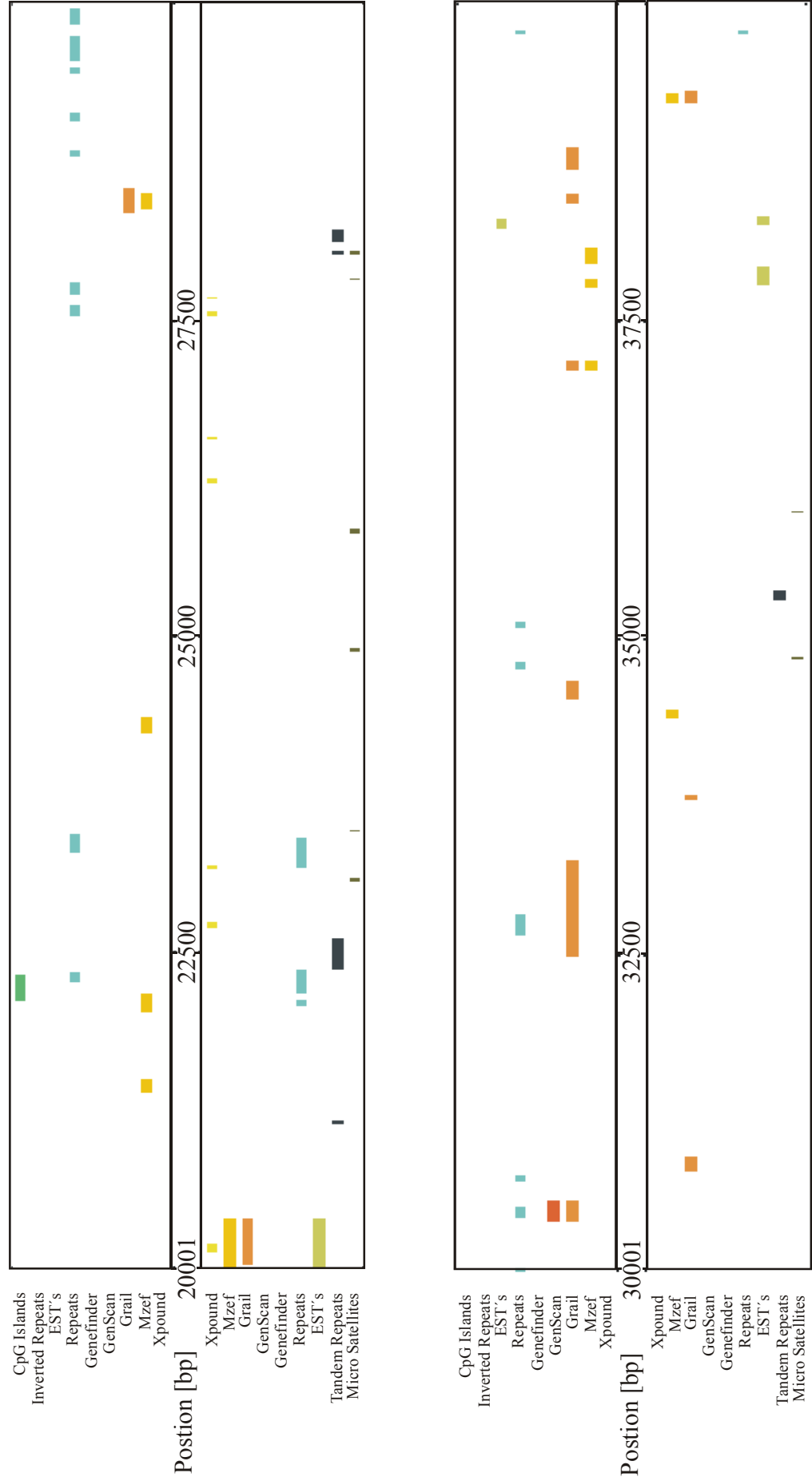
FT          /note="94% identity: matches 423..569 of consensus"
FT  repeat_region complement(113778..113924)
FT          /rpt_family="L1"
FT          /note="89% identity: matches 161..307 of consensus"
FT  repeat_region 113927..113961
FT          /rpt_type=TANDEM
FT          /rpt_family="aaaaaat repeat"
FT          /note="homology = 91.4%, counts = 5"
FT  repeat_region 114871..115141
FT          /rpt_type=INVERTED
FT          /note="IR16', 86% complementary to IR16 (113681..113953)"
FT  repeat_region 114898..115123
FT          /rpt_family="L1"
FT          /note="84% identity: matches 245..468 of consensus"
FT  repeat_region complement(114898..115141)
FT          /rpt_family="ALU"
FT          /note="89% identity: matches 41..281 of consensus"
FT  repeat_region complement(115386..115550)
FT          /rpt_family="THE"
FT          /note="87% identity: matches 55..222 of consensus"
FT  repeat_region complement(115562..115753)
FT          /rpt_family="THE"
FT          /note="86% identity: matches 1..195 of consensus"
XX
SQ  Sequence 116189 BP; 38824 A; 21112 C; 20922 G; 35331 T; 0 other;

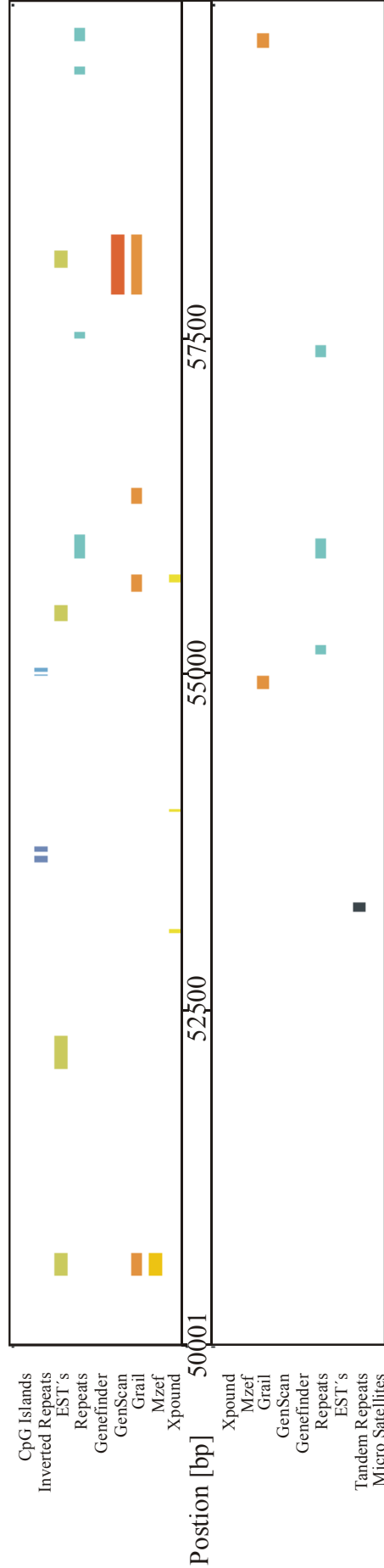
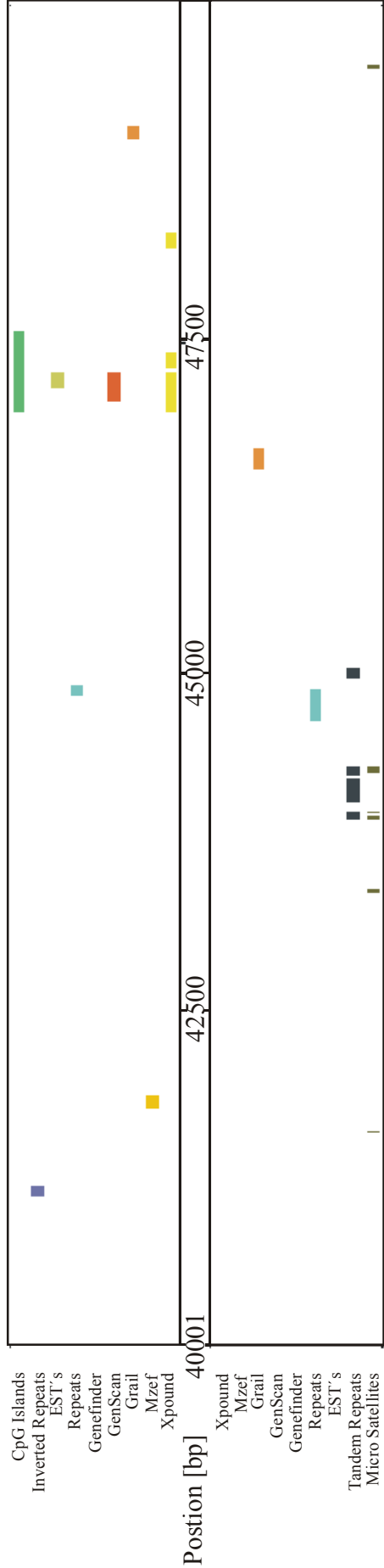
```

7.4.1 Graphische Darstellung der Annotierungsergebnisse

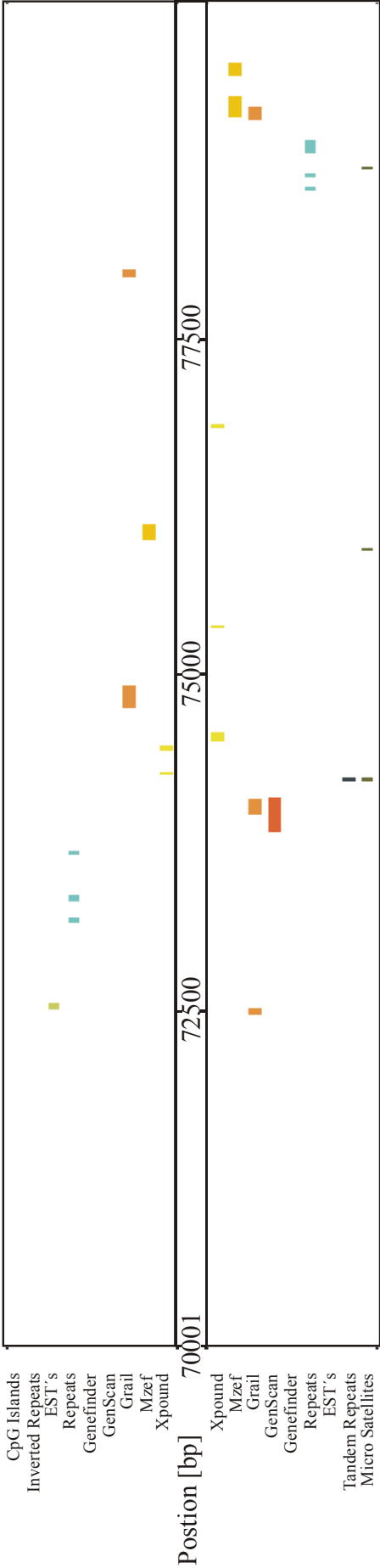
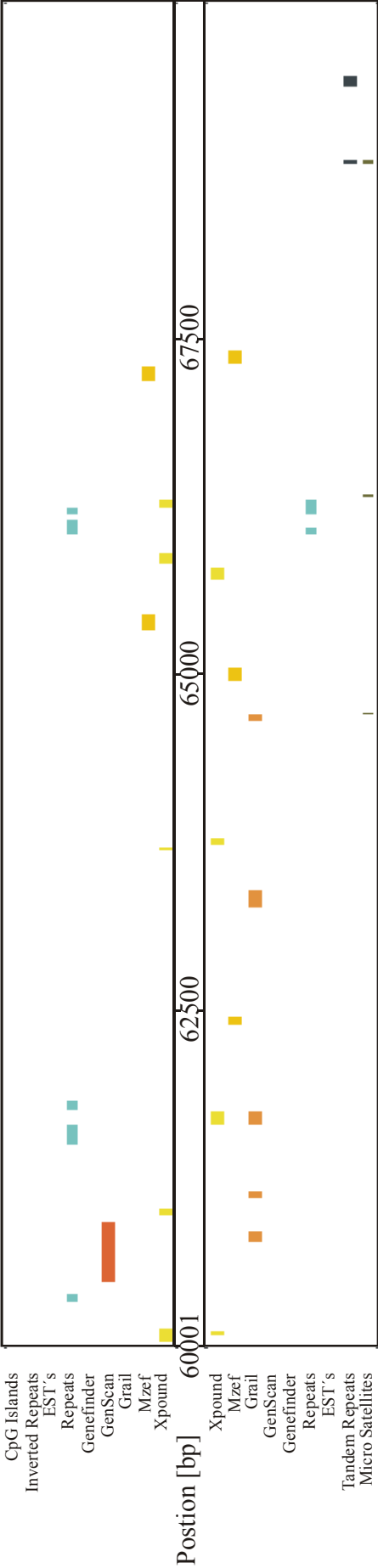
Als Beispiel sind die graphischen Annotierungsdaten für den Klon 98L15 von Chromosom 21 gezeigt, die aus den textbasierten Annotierungsdaten (Kapitel 7.4) dieses Klons mit dem Programm „showseq“ erzeugt wurden.

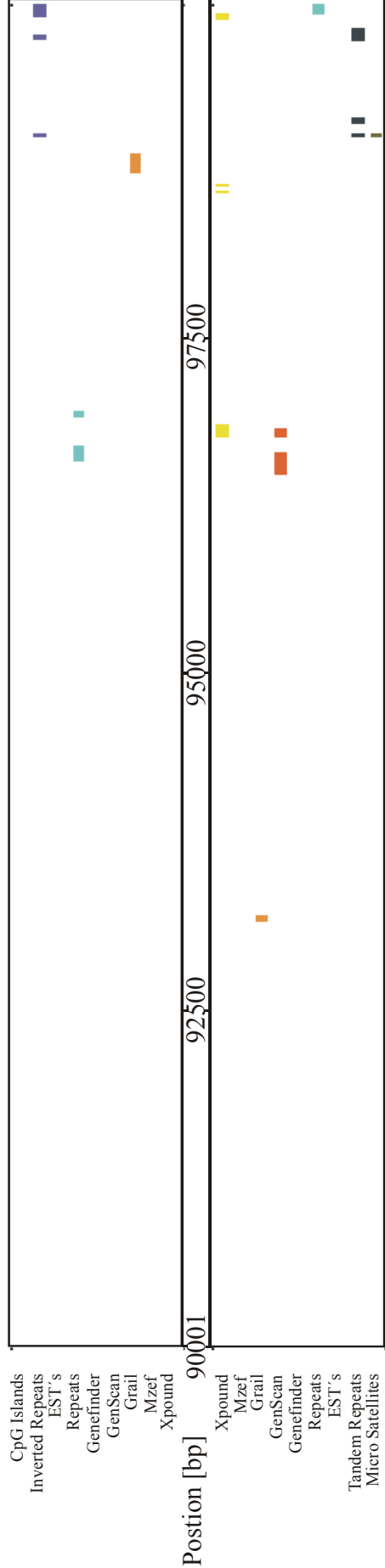
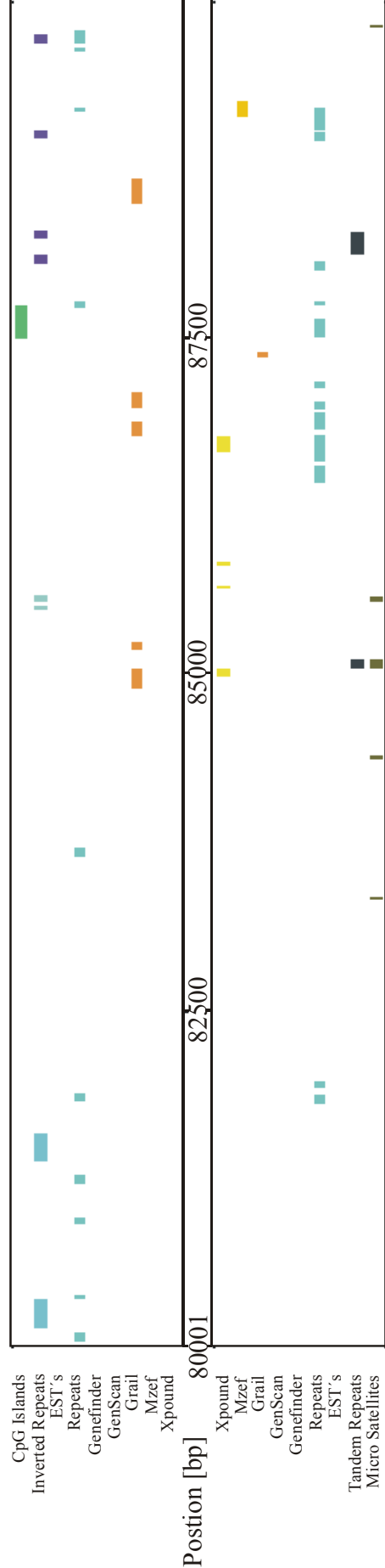


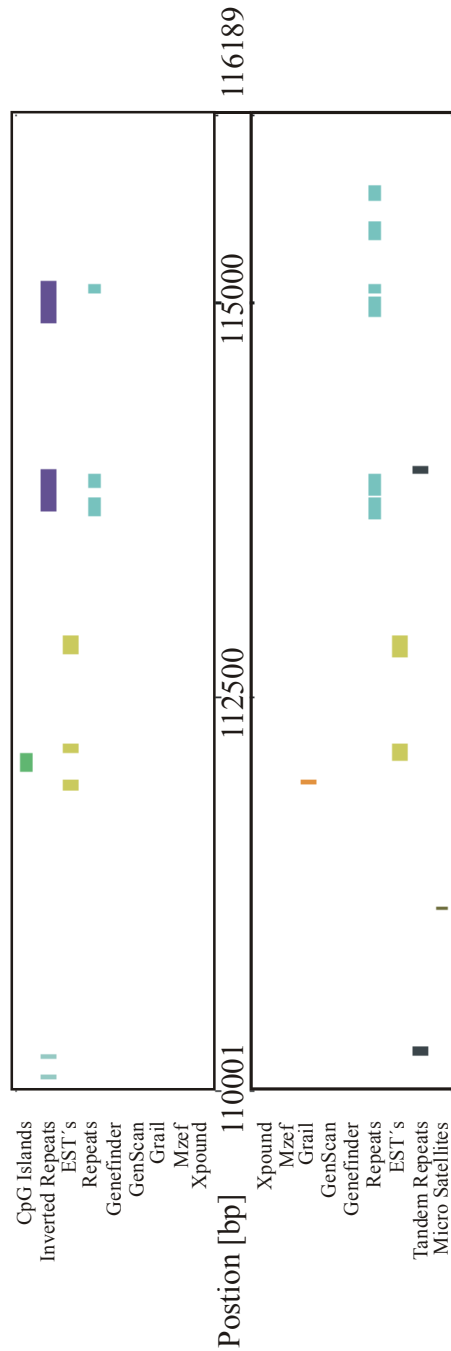
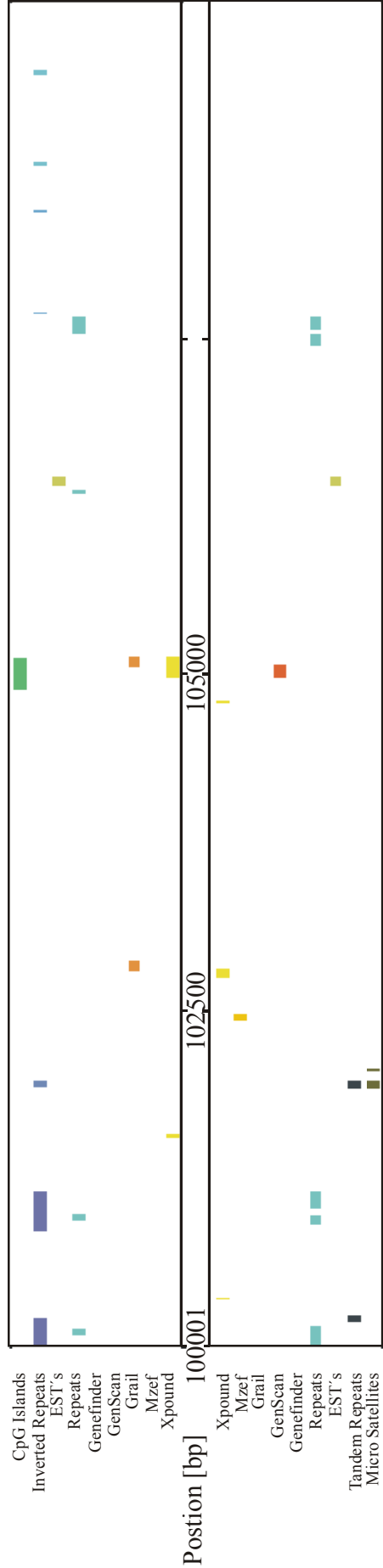




Anhang







Danksagung

Die vorliegende Arbeit wurde in der Abteilung Genomanalyse an der Gesellschaft für Biotechnologische Forschung mbH in Braunschweig angefertigt.

Allen voran möchte ich meiner Familie für die liebe Hilfe und Unterstützung während dieser Zeit herzlich danken.

Mein besonderer Dank gilt Herrn Prof. Dr. J. Bode und Herrn Prof. Dr. K. Buchholz für das Interesse an meiner Arbeit und für die Übernahme des Haupt- bzw. Korreferats. Bei Herrn Prof. Dr. H. Schmidt bedanke ich mich für die Teilnahme an der Prüfungskommission.

Mein Dank gilt Herrn Dr. H. Blöcker für die Überlassung des Themas und der gewährten Freiheit bei der Gestaltung meiner Arbeit.

Bei den Mitarbeiterinnen und Mitarbeitern der Abteilung Genomanalyse bedanke ich mich für das angenehme Arbeitsklima. Frau Dr. P. Brandt und Dr. G. Nordsiek danke ich für die Unterstützung und Diskussionsbereitschaft bei der DNA-Sequenzanalyse. Herrn Dr. M. Czubyko danke ich für die netten Diskussionen, die auch mal über den fachlichen Bereich hinaus gingen. Weiterhin möchte ich mich für die kritische Durchsicht des Manuskripts bei Herrn Dr. G. Kauer und H. Stiege bedanken.